DISCUSSION PAPER / 2023.01

Effective Altruism and the strategic ambiguity of 'doing good'

Mollie Gleiberman

University of Antwerp
IOB | Institute of Development Policy

Comments on this Discussion Paper are invited.

The IOB Discussion Papers (ISSN 2294-8651) are peer reviewed. Publication as an IOB Discussion Paper does not constitute prior publication, and does not preclude publication elsewhere. The findings and views expressed in the IOB discussion papers are those of the authors. They do not necessarily represent the views of the IOB as an institute.

Instituut voor Ontwikkelingsbeleid en -Beheer
Institute of Development Policy

Postal address:
Prinsstraat 13
B-2000 Antwerpen
Belgium

Visiting address:
Lange Sint-Annastraat 7
B-2000 Antwerpen
Belgium

Tel: +32 (0)3 265 57 70
Fax: +32 (0)3 265 57 71
e-mail: iob@uantwerp.be
http://www.uantwerp.be/iob

DISCUSSION PAPER / 2023.01

Effective Altruism and the strategic ambiguity of 'doing good'

Mollie Gleiberman*

April 2023

*PhD student at IOB. Contact: dmgleiberman@gmail.com

Acknowledgments: I am grateful to two anonymous reviewers and to Sara Geenen for thoughtful comments and constructive feedback on an earlier version of this paper; all remaining errors are my own.

Table of contents

Abstract

Abstract: This paper presents some of the initial empirical findings from a larger forthcoming study about Effective Altruism (EA). The purpose of presenting these findings disarticulated from the main study is to address a common misunderstanding in the public and academic consciousness about EA, recently pushed to the fore with the publication of EA movement co-founder Will MacAskill's latest book, *What We Owe the Future* (WWOTF). Most people in the general public, media, and academia believe EA focuses on reducing global poverty through effective giving, and are struggling to understand EA's seemingly sudden embrace of 'longtermism', futurism, artificial intelligence (AI), biotechnology, and 'x-risk' reduction. However, this agenda has been present in EA since its inception, where it was hidden in plain sight. From the very beginning, EA discourse operated on two levels, one for the general public and new recruits (focused on global poverty) and one for the core EA community (focused on the transhumanist agenda articulated by Nick Bostrom, Eliezer Yudkowsky, and others, centered on AI-safety/x-risk, now lumped under the banner of 'longtermism'). The article's aim is narrowly focused on presenting rich qualitative data to make legible the distinction between *public-facing* EA and *core* EA.

1. Introduction

Over its ~10 year history, Effective Altruism (EA) has typically been (mis)understood[1] by academics, the general public, and casual participants as a philosophy and social movement focused on encouraging philanthropic donations to evidence-based aid interventions and charities serving the global poor. This perception treats EA as synonymous with the practice of encouraging people to concentrate their donations on charitable interventions backed by rigorous, quantitative evidence such as randomized controlled trials (RCTs). EA has built a formidable reputation based on this understanding, wherein Effective Altruists (EAs) are depicted as 'charity nerds' (Piper, 2016) willing to dramatically reduce their personal spending in order to donate more money to charities proven to be highly *effective* in terms of maximizing the number of lives saved or improved per dollar (Burton, 2015; Matthews, 2013; Ord, 2011), such as those providing anti-malarial bednets, deworming pills, and cash grants to people in sub-Saharan Africa (reflecting the top recommendations of GiveWell, EA's flagship charity-evaluator).

However, recent developments, such as the publication of movement co-founder Will MacAskill's new book *What We Owe the Future* (hereafter: WWOTF), have generated considerable confusion in the public perception of EA, namely, EA's seemingly-new embrace of 'longtermism' and the prioritization of research and policymaking related to the development of safe artificial intelligence (AI-safety/AI-alignment) and the reduction of existential risks ('x-risks') to the future of humanity. This confusion exists because many people are unaware that EA has always been intensely focused on the agenda that is now advanced under the banner of 'longtermism' particularly AI-safety and x-risk/global catastrophic risk reduction[2] . For instance, in 2012, the year the movement formally launched, MacAskill wrote approvingly that 'new effective altruists tend to start off concerned about global poverty or animal suffering and then hear, take seriously, and often are convinced by the arguments for existential risk mitigation' (Crouch [MacAskill], 2012c). By 2016, the main funder of the EA movement, Open Philanthropy (an organization spun-off from GiveWell which disburses the fortune of Facebook co-founder Dustin Moskovitz) had designated AI-safety a priority area (Karnofsky, 2016b, 2016c, 2016d, 2016e). AI/x-risk has been prominently featured in leading EA organizations' material

[1] The movement's founders and leaders are the first to acknowledge that this misunderstanding exists: 'if I had to sum up the misunderstandings with effective altruism in one line, it's just that people think that effective altruism is just about the claim that we should donate money to evidence-backed interventions that help the world's poorest people' (Ben Todd, in Koehler et al., 2020); 'Effective altruism is widely misunderstood, even among its supporters [...] In short, effective altruism is commonly viewed as being about the moral obligation to donate as much money as possible to evidence-backed global poverty charities, or other measurable ways of making a short-term impact' (Todd, 2020a); '[The Centre for Effective Altruism's] understanding of effective altruism, which is widely endorsed by those within the effective altruism community, is quite different from the understanding of effective altruism possessed by many in the general public and by many critics of effective altruism' (MacAskill, 2019, p. 11).

[2] For some examples see (80000 Hours, 2012; Beckstead, 2011a, 2011b, 2013; Christiano, 2010, 2013; Crouch [MacAskill], 2012d; Mindermann, 2013; Nachbauer, 2011; Piper, 2014; Redwood, 2011, 2012a, 2012b; Redwood & Muehlhauser, 2012; Todd, 2013a, 2014; User: Raemon [Raymond Arnold], 2012a, 2012b)

for many years (Centre for Effective Altruism, 2018a; Whittlestone, 2015; Wiblin, 2015b), with AI/x-risk consistently topping the list of what leading EA organization 80,000 Hours considers to be the world's most pressing problems since 2016 (see, e.g., 80000 Hours, 2016a; 80000 Hours, 2016b, 2017, 2021, 2022a).

Given the evidence of EA's longstanding commitment to prioritizing AI-safety/x-risk and 'longtermism', why do so many people continue to insist that EA is *really* just about evidence-based aid interventions serving the global poor? Why, with the recent publication of MacAskill's *What We Owe the Future*, are journalists and academics scrambling to account for EA's seemingly sudden interest in AI, treating 'longtermism' as a deviation from *authentic* EA? This article aims to help answer that question, by analyzing how EA rhetoric operates across two registers. It argues that the EA movement has historically produced two types of content aimed at different audiences: *public-facing* EA content designed to attract a general audience and broad support, published in popular media and books targeting a general audience, and *core* EA content that reflects the ideas and aims of the movement's leaders and highly-engaged members, and which is discussed on inward-facing spaces, particularly the online EA community forums. The strategic ambiguity of phrases like 'doing good', 'helping others', and 'caring about the world' enables these two projects to operate under the same name, with one project (the Singer-inspired approach to *effective giving*) employed to establish the movement's credibility and encourage new recruits, some of whom are then further recruited to support an entirely different project (core EA, which reflects the transhumanist agenda articulated by Nick Bostrom, Eliezer Yudkowsky, David Pearce, Aubrey de Grey, Robin Hanson and others in the early 2000s,[3] now rebranded as 'longtermism').

The article is primarily descriptive, and aims to portray a narrow but deep sliver of some initial empirical findings from a larger study about EA. Thus the article does *not* engage with existing academic debates about the role of EA in global health and development efforts, not because such debates are unimportant or uninteresting, but because—as movement leaders and core members themselves admit—EA is not primarily concerned with evaluating aid effectiveness or encouraging donations to alleviate global poverty.

Following this introduction and a discussion of the methodology (section 1), the paper is divided into four main sections:

- In **section 2**, I offer a **brief overview of the EA movement** (history, key ideas, major institutions/figures) to help orient readers unfamiliar with EA.
- In **section 3**, I look at **how EA's double meanings are *constructed***. I first outline several concepts that help illuminate this strategy of double-meanings—tactical polyvalence, glittering generalities, and frame-appropriation—and then show how this strategy leads to two different versions of EA existing simultaneously (public-facing EA and core EA).
- In **section 4**, I look at **how EA's double meanings are applied** in outreach to different audiences, and how the use of identical terms to refer to different things acts

[3] Transhumanism is a techno-futurist ideology mixed with Silicon Valley libertarianism, characterized by advocacy of artificial intelligence (AI), prediction and forecasting markets, belief in the coming of the Singularity (the moment at which artificial intelligence becomes equal to or surpasses human intelligence, i.e. the creation of greater-than-human intelligence, also known as 'superintelligence' and 'artificial general intelligence' [AGI]); and the goals of colonizing space, genetically enhancing the moral/physical/cognitive capacities of both human and non-human animals, and eliminating death via cryonics, whole brain emulation (uploading human consciousness into supercomputers to create digital minds) or other life-extension biotechnologies (Bostrom, 1999, 2005a; Sandberg, 2015). In the early 2000s, transhumanists like Eliezer Yudkowsky, Aubrey de Grey, and Nick Bostrom (who is transhumanism's most prolific writer) initiated an effort to turn transhumanism—till then considered the realm of internet cranks and science fiction junkies—into 'a more academically respectable and intellectually serious inquiry' (Bostrom, in Humphrey, 2004). A key aspect this strategy was framing advocacy *for* emerging technologies like AI/AGI and biotechnology encased within concerns about their *safe* development and the prevention of existential risks ('x-risks') that would cause human extinction or the curtailment of the technological progress that transhumanists desire to bring about (Bostrom, 2002, 2003, 2005b). The transhumanists specifically invoked the use of utilitarian calculations to classify transhumanist goals as the optimal use of philanthropic donations, in terms of saving or improving the most lives per dollar donated (see, e.g., Howe, 2002; Singularity Institute for Artificial Intelligence, 2002).

as a credibility bridge that leads recruits from one version of EA into adopting the other version of EA. This section focuses on the 'funnel model' of EA outreach and recruitment.

- In **section 5**, I examine the **tensions that arise from these double-meanings**, particularly in the present moment where the glaring discrepancy between the public's longstanding beliefs about EA and the actual practices of the EA movement has become obvious and undeniable.

1.1. Methodology

The article presents findings from an interpretive, qualitative study grounded in an inductive, ethnographic approach, wherein my 'field' was the technologically-mediated landscape of online spaces (Hine, 2017). Findings stem from four years as a non-participant observer immersed in the online EA community, complemented by extensive collection of online documentary artifacts throughout that period (and ongoing into the present). I assembled a master database of over 12,000 text-based artifacts collected via *unobtrusive* (Hookway & Snee, 2017) or *passive* (Keim-Malpass et al., 2014; Wimmer & Dominick, 2014) *online research*, i.e., accessing content produced and posted publicly to the internet of user's own volition, without any involvement by the researcher. This approach is particularly well-suited to EA: EA evolved out of an online community that gathered around forums/blogs like *Felicifia*, *LessWrong*, *Slate Star Codex*, and *Overcoming Bias*, and EAs continue to use online forums (especially the *EA Forum*) to discuss ideas, strategies, and the direction of the movement[4]. While I occasionally reached out to EA organizations for clarifications on specific information[5], I did not attempt to engage with the community beyond this; given the high level of control leading organizations like the Centre for Effective Altruism (CEA) exercise over how EA is presented to outsiders[6], efforts to interview movement members would likely yield responses far less candid than the discussions which can be observed in EA's publicly-available online spaces[7]. I did not create user accounts on the forums I read, nor did I attempt to access private EA spaces, such as non-public/password-protected Slack/Discord channels.

Primary sources were collected iteratively, by continuously comparing new data against previously-collected data (Bowen, 2008; Yanow, 2006), allowing themes and patterns to emerge (Bowen, 2009). Analysis employed mixed-methods, combining three complementary strategies: *thematic content analysis* (close reading of texts), *website history tracing* (the archaeology of digital content using online archives to trace the evolution of web site content over a period of years, and to uncover deleted web pages[8]), and *funding-pattern analysis* (a 'follow-the-money'

[4] EAs themselves acknowledge that the online EA culture is a crucial component of the movement, and that the *EA Forum* is a central node of the social movement (Dalton, 2018b).
[5] I also discovered that basic fact-checking requests on certain topics simply went unanswered: in October 2022, I made several requests to CEA for confirmation that they had purchased Wytham Abbey for ~£15m and the Lakeside Guesthouse for ~£1.5m. I received a reply asking me why I needed this information, and when I explained why I thought the purchases were of public/academic interest, CEA/EV did not reply to that email nor a follow-up email sent later. Eventually, I obtained copies of the title deeds for the properties directly from the UK's public land registry.
[6] The movement has published numerous official and informal guides that carefully instruct members how to talk about EA in different contexts and to different audiences, including interviews with journalists and researchers, and CEA advises all community members to consult with CEA for advice, guidance, and coaching before granting interviews (Mayhew, 2019; Wise et al., 2020). See, for example, "*A Guide to Early-Stage EA Group-Building at Liberal Arts Colleges*" (Agarwalla, 2019); "*The EA Pitch Guide*" (Alterman, n.d.); "*Tips to Help Your Conversation Go Well*" (EA Hub, 2019a); "*What to Say*" (EA Hub, 2019c); "*Talking About Effective Altruism*" (Effective Altruism Wiki, 2015); "*A Model of an EA Group*" (EffectiveAltruism.org, 2019); "*Heuristics from Running Harvard and Oxford EA Groups*" (Flidr & Aung, 2018); "*Guide to talking about effective altruism and effective giving*" (Freeman et al., 2021); "*Talking About Effective Altruism at Parties*" (Gertler, 2017); "*Talking About EA*" (Gertler, n.d.); "*Starter Packet For Giving What We Can Chapters*" (Giving What We Can, 2012); "*Workplace Activism: EA Handbook*" (Giving What We Can, 2016); "*Effective Altruism: Chapter Starter Packet*" (Giving What We Can, n.d.); "*Altruism Icebreakers*" (Sittler [Adamczewski], 2015); "*What to Know Before Talking to Journalists about EA*" (Mayhew, 2019); "*A Vision for Harvard University EA Student Group (HUEASG) 2018-19*" (O'Keefe, 2018); "*How We Promoted EA at a Large Tech Company*" (User: ParthThaya, 2020); "*Advice for Responding to Journalists*" (Wise et al., 2020); "*EA Student Group Handbook – Packaged Programs*" (Global Challenges Project, 2021a).
[7] The *EA Forum* and other online spaces are also important as places where EA is not merely discussed, but *performed*; interpretive analysis requires cultivating sensitivity to what a text intends to convey to its audience (regardless of its facticity), i.e., what meanings are being actively constructed.
[8] While I will not delve into the ethics of online research in the present article, the recognition of the need to act

exercise tracking financial flows across the EA ecosystem). I use extensive quotes and references to this primary data to substantiate my claims, all of which can be verified for their accuracy and context via links provided in the reference list.

The picture of EA presented here emerges from analysis of 4,836 text-based artifacts selected from the master database. The artifacts were produced by EA organizations, leaders and members, funders, or affiliated organizations/movements, and published publicly online. Of these, 1,197 come from the three key community forums: *EA Forum* (n=849), *Felicifia* (n=106; defunct since 2014 but important for understanding EA's historical development), and *LessWrong* (n=242)[9]. These sources were located using an internet search engine to identify major references about EA, then snowballing outwards from these to identify key individuals, blogs, forums, and other organizations. Data collection also involved using web archives (e.g., Internet Archive's Wayback Machine) to trace the evolution of organizational web pages over time; this enabled me to monitor changes in mission statements and vocabularies, track the movement of key personnel between and across organizations, and to recover deleted or expired pages relevant to the study. Primary data was collected from the online content of over 233 EA/EA-affiliated (including 'rationalist' and 'transhumanist') organizations (including general web pages, blogs, public social media accounts, white papers, strategic plans, conference schedules, handbooks, style guides, research agendas, audio recordings of meetings, videos of workshops/talks/lectures, press releases, annual tax forms and audited financial statements); posts and comments on community forums; the individual websites or blogs of approximately 90 movement leaders/thought-leaders and influential core members[10]; course syllabi; grant award announcements, requests for proposals, and grants databases of EA funders; popular media produced by movement members aimed at the general public; conference presentations; videos and podcasts. The majority of this primary data was originally produced between 2006 and 2022, although some dates from as early as 1996 and as late as February 2023 (during revisions to this article). By ensuring that collection covered a long period of time, I was able observe patterns of communication over a period of years, facilitating the 'thick description' that online ethnography seeks to convey (Janetzko, 2017).

Primary Sources vs. Secondary Sources

The article is descriptive, and most of the cited references are *primary sources* (pieces of data that provide the evidence). It is a common convention in qualitative research to provide a reference to primary data (e.g., an interview, primary document, etc.) in order to indicate the source of evidence for the claim being made. When dealing solely with documentary artifacts, it can be confusing for readers to differentiate between references to primary sources versus a secondary sources, since the visual representation —"(Name, Year)"— is identical. I have separated out "Primary Sources" in the reference list for the readers' convenience, along with other kinds of primary data e.g., government documents from the UK and US government land registries.

The line between primary and secondary sources can be blurry, and the same artifact (piece of text) may be categorized as both depending what information is being gathered from it. For example, journalist Gideon Lewis-Kraus' profile of Will MacAskill for the *New Yorker* is both a primary source for my study (it provides a data point on how journalists present EA to readers) *and* a secondary source (it presents facts and data gathered by Lewis-Kraus *about* the EA movement). Likewise, most of the peer-reviewed academic literature about EA that I refer-

with ethical integrity regarding online content (particularly that which users might have intentionally deleted) was included in considerations about data collection strategies.

[9] For this study, I have prioritized posts/comments by the movement leaders and core movement members, posts that are marked by site moderators as 'front page' or 'curated' (indicating that they are recommended reading), posts that deal specifically with communicating about EA, community-building/outreach, and debates about core EA ideas. I also read the most upvoted/recommended posts each week to gauge what topics are 'trending' within the community (while also looking at downvoted posts/comments, which provide information about topics/debates the community does not approve of), while also following links embedded within posts so as to find relevant community-discussions that I might otherwise miss. I continue to rely on the *EA Forum* for regular updates, e.g., new EA organizations, announcements, current events, and linkposts for important EA content posted elsewhere.

[10] See Appendix A for a representative list of community forums and blogs consulted.

ence are *actually* primary sources here, because I am not using these articles to gain insights into arguments or to learn facts about EA, but as data points illustrating how academics present EA to readers. However, in an effort to clearly distinguish these sources from the sources produced by the EA community and sources more traditionally understood as 'primary' data, I have classified these types of sources as "Secondary Sources" in the reference list.[11]

Usernames vs. Real Names

Generally, 'doxing' (revealing the real identity of someone who posts pseudonymously on the internet) is frowned upon in internet culture (Douglas, 2016), requiring a careful approach regarding when to connect a username to a real identity. In the forums I viewed, individuals typically tie usernames to real identities. I attribute forum comments/posts *only* to usernames when an author is not publicly known by both (e.g. a quote from a hypothetical 'Charles24' who does not openly link their username to their real name will be referenced as 'User: Charles24'). In cases where a user's real identity is publicly linked but may not be obvious to the reader, I indicate both the username and real name in the citation. In cases where a username is virtually identical to a real name, I cite real name only. Importantly, some highly-influential community-members only post under usernames. The fact that someone writes pseudonymously under a username does not indicate peripheral involvement or low status in the community; on the contrary, some extremely influential community members (e.g., Utilitymonster on *Felicifia*, Gwern on *LessWrong*, and Scott Alexander of *Slate Star Codex*) do not post under their real names.

Who is an 'Effective Altruist'?

Some people whose admiration for or participation in EA is based on public-facing content (i.e., people who believe EA is a movement for effective giving started by Peter Singer, and treat the phrase 'effective altruism' as nothing more than a philosophical idea) may instinctively reject the portrayal here, since, for them, EA *really is* a genuine effort to reduce global poverty[12]. These individuals might protest that since *they* consider themselves 'effective altruists', and *they* privilege global poverty reduction, then 'effective altruism' *is* about addressing global poverty. I refer to such individuals as grassroots effective altruists (for lack of a better term to distinguish them from 'EAs', that is, the people for whom participation in the EA movement is a core part of their professional and/or personal identity). Grassroots effective altruists tend to believe that EA is a philosophy advocated by a loose, decentralized set of organizations with a shared mission, not realizing that it is a tightly coordinated and hierarchical social movement that carefully protects its brand. It is true that all social movements and organizational cultures are multifaceted, riven with internal debates and disagreements (Snow, 2004); EA is no exception. Of course, *all* EAs and people who identify with or follow the EA philosophy do not uniformly hold *all* the views which I ascribe to EA *qua* the EA movement—even amongst the leadership, there are many disagreements. Yet it is possible to acknowledge that heterogeneity, divergence, and conflict exist in *any* movement, while still recognizing that in-group members share certain overarching goals, values, and epistemic frameworks, else they would not consider themselves members of the movement (Van Dijk, 2006). No group speaks with a single voice, but members of social movements do have a *collective identity* — a sense of being a 'we', regardless of internal discord — that is *relational* in terms of what is shared amongst members (insiders, 'us'), against the external environment to which they stand in contrast (outsiders, 'not-us') (Melucci, 1995; Smithey, 2009; Snow, 2001). Grassroots effective altruists occupy the liminal space between the 'we' and the 'not we' of the EA collective identity.

At the same time, certain perspectives continuously triumph over others in contestations over framing and priorities in EA: the perspectives and priorities of a core inner circle of leadership figures (including thought-leaders whose influence remains relatively behind-the-scenes), who founded EA, direct the major EA organizations, and hold considerable decision-making

[11] This is an imperfect system, but intended to address an important point raised by an anonymous reviewer. I am grateful to this reviewer for pointing out the necessity of clarifying this ambiguity.
[12] For simplicity, here I am bracketing EA's less-emphasized but also public-facing content on ending factory farming, which is another facet of public-facing EA. The rhetoric for which EA became famous focused mainly on global poverty examples.

power, particularly through their influence over funding decisions. Thus, while the goals of EA may be contested and variable, this does not mean that one cannot discern a version of EA that *more accurately* represents the goals of the movement's founders, highly-engaged members, and funders[13], who influence the direction the movement takes and *without whom the movement simply would not exist*. Supporting this contention is the high level of conformity that exists in the statements by major EA organizations, leaders and core members, and is backed up by the decisions of EA funders and grant-makers regarding *who* and *what activities* to fund.

This article presents a version of EA that some readers might find disheartening, particularly those who subscribe to the version of EA that I call 'grassroots effective altruism'. It is important to acknowledge that many people who embrace the philosophy of public-facing EA do admirable work based on that understanding, including making considerable sacrifices so as to donate generously to the needy. There are also some people deeply involved in the EA community, who self-identify as EAs, and are fully aware of EA's increasingly predominant focus on AI/x-risk/'longtermism', but who continue to prioritize 'neartermist' causes. It is also clear that EAs who embrace 'longtermism' do so out of genuine conviction that this is the optimal way to help others — regardless of how misguided it may seem to outsiders. In other words, I have no doubt that *everyone* working under the banner of EA *sincerely* believes that they are *doing good*. This belief, however, does not preclude the possibility that some efforts to do good might in fact be doing harm, that people pursuing good intentions may go astray, or that one group's idea of what is 'good' for everyone might be another group's nightmare. Indeed, for a movement that is oriented around *consequentialism*—the position that outcomes, not intentions, are what truly matter— the possibility that *stated good intentions* might be leading to *potentially harmful outcomes* seems like something EA movement sympathizers would want to shine a light on.

2. Background

2.1. A Brief History of EA

The origins of the EA movement are generally traced to two organizations with similar missions, both founded in 2007: **GiveWell** and **Giving What We Can (GWWC)**. GiveWell was established in New York City by Holden Karnofsky and Elie Hassenfeld (GiveWell, 2007a, 2007b, 2007c), and GWWC was informally launched on a utilitarian forum called *Felicifia* by a philosophy graduate student at Oxford named Toby Ord (Giving What We Can, 2007; Ord, 2007); GWWC later formally relaunched in 2009 as a joint-effort by Ord and another philosophy grad student named Will Crouch (now MacAskill). Both organizations ranked charities according to a utilitarian-inspired cost-effectiveness calculation of *most lives saved or improved per dollar*. GWWC also introduced a **'giving pledge'** inspired by utilitarian ethicist Peter Singer—an early and vocal supporter of both organizations, who is often mistakenly portrayed as a founder of the EA movement due to his support—wherein GWWC members committed to 'give at least ten percent of what I earn to whichever organizations can most effectively use it to fight poverty in developing countries' (Giving What We Can, 2011). In 2010, the first US chapter of GWWC was launched at Rutgers University by three philosophy students: Nick Beckstead, Mark Lee, and Tim Campbell (Heyboer, 2010; Roache, 2010; Yetter Chappell, 2010).

Soon the individuals affiliated with GiveWell and GWWC were interacting on the *GiveWell blog* and various online forums dedicated to utilitarianism and rationality, such as *Felicifia*, *LessWrong*, and *Overcoming Bias*. In 2011, MacAskill and another Oxford student, Ben Todd, launched a career advisory organization called **80,000 Hours (80K),** representing the average number of hours someone works over their lifetime. 80K aimed to help students find high-

[13] In ~2016/17 the emphasis on AI-safety/x-risk became more openly endorsed by leadership. Community-members (and some leaders) who disagreed began to voice concerns that CEA/80K/GWWC were dictating EA priorities in ways that contradicted what the grassroots effective altruist community prioritized (see the various comments on Dalton, 2018a). Such discussions presume that EA is a grassroots movement with no clear center of gravity—essentially, that what EA *is* remains up-for-grabs —when in fact EA was created by a small group of individuals holding a particular vision, with the goal of pursuing that vision. EA is, essentially, what the people who created it and who lead major EA organizations say it is (Todd, 2020b). Since 2018, many of the individuals who at first resisted the 'new' focus on AI/x-risk/longtermism seem to have drifted away from the movement, but still embrace the grassroots effective altruist vision (Melchin, 2020).

impact careers—the original name was **High Impact Careers (HIC)**, changed last minute to the catchier-sounding '80,000 Hours' after a discussion on *Felicifia* (User: Arepo [Sasha Cooper], 2011)—based on the idea of 'earning to give'[14]: taking an especially high-paying job in a sector not traditionally seen as altruistic (e.g. banking/finance, tech) and then donating a high proportion of one's income to charity, rather than working directly in a non-profit (Crouch [MacAskill], 2011; Crouch [MacAskill] & Todd, 2011; MacAskill, 2013f, 2014c)[15]. Shortly after 80K launched, EA formally crystallized as a social movement with the launch of the **Centre for Effective Altruism (CEA)** in February 2012 by Toby Ord, Will MacAskill, Nick Beckstead and Michelle Hutchinson (Companies House - GOV.UK, 2012). Both GWWC and 80K were brought under CEA's umbrella and the movement became officially known as 'effective altruism' (Reese, 2015). New, interlinked organizations operating under the aegis of EA sprang up, such as another charity evaluator called **The Life You Can Save (TLYCS**; later spun off under the direction of Peter Singer), **Animal Charity Evaluators (ACE)**, **Leverage Research**, and a project of Leverage called **The High Impact Network (THINK).**

Meanwhile, GiveWell was evolving: after establishing a reputation for *only* recommending 'proven, effective, and scalable' interventions (Karnofsky, 2007), the organization—flush with the support of Facebook co-founder Dustin Moskovitz and his wife Cari Tuna, the billionaire couple behind **Good Ventures**—launched a project called **GiveWell Labs** in September 2011 (Karnofsky, 2011) to explore interventions that were unproven (thus risky), but could lead to exceptionally high social impact *if* successful—what they later came to refer to as 'hits-based giving' (Karnofsky, 2016a). In 2014, GiveWell Labs rebranded as the **Open Philanthropy Project** (now just 'Open Philanthropy', **OpenPhil**), eventually becoming a separate organization in 2017 with Karnofsky as CEO and Tuna as President (Karnofsky, 2017).

2.2. Current EA

EA is typically explained as both a philosophy and a social movement. In theory, EA seems to be little more than an ethical injunction to do the most good one can with the resources (time, money, skill, etc.) that one has at their disposal. In practice, however, the EA movement—in other words, *actually-existing* EA—is a vast network comprising hundreds of organizations[16] but led by a relatively small group of influential leadership figures at a handful of key organizations. In addition to those mentioned above, other key organizations include: the **Center for Applied Rationality (CFAR)**, **Rethink Priorities**, **Longview Philanthropy** (formerly Effective Giving UK), the **Global Priorities Institute (GPI)**, **Forethought Foundation for Global Priorities (Forethought)**, the **Center on Long-Term Risk (CLR**, formerly the **EA Foundation**), **Founders Pledge**, the **Machine Intelligence Research Institute (MIRI)** and the **Future of Humanity Institute (FHI)**. These organizations provide intellectual capital to the EA ecosystem, and have nurtured the further development of hundreds of other EA organizations, many of which are funded primarily by OpenPhil, either directly, or indirectly through OpenPhil's support to re-granting bodies like CEA's **EA Funds** (specifically, the **Long Term Future Fund [LTFF]**, and the **EA Infrastructure Fund [EAIF])**, the **CLR Fund**, and the **Berkeley Existential Risk Initiative (BERI)**. Another major EA funder of note is Skype co-founder Jaan Tallinn (Tallinn, 2020), who makes grants directly through his **Survival and Flourishing Fund** and indirectly through his donations to BERI and CEA's EAIF and LTFF.

[14] Originally an idea posited by Peter Unger in *Living High and Letting Die* (Unger, 1996), the idea of earning to give is generally credited with entering the *Felicifia* community (Kaufman, 2012) via prominent Felicifian Brian Tomasik's 2006 essay "Why Activists Should Consider Making Lots of Money" (Tomasik, 2006) posted on his blog, *Utilitarian Essays*.
[15] 80K no longer advocates earning to give, but instead offers one-on-one coaching and advice to students seeking careers in areas that 80K designates as 'priority paths'. Note that 80K's 'top-recommended career paths' center on technical AI research and AI policy/governance, biorisk/biosecurity, building the EA movement, and prediction and forecasting related to x-risks/emerging technologies (80000 Hours, 2022b).
[16] See table in Appendix B.

CEA remains the guardian of the EA brand and direction. Outsiders and casual movement members imagine EA as a loose coalition of organizations sharing a mission; however, in practice, CEA and those closely affiliated with it set priorities and steer the movement: CEA maintains the ***EA Forum*** (the main online forum); publishes the ***EA Handbook***; maintains **EffectiveAltruism.org**; coordinates and funds hundreds of local EA chapters at universities around the world; organizes and selects attendees for the major EA conferences each year, **EA Global**; organizes the annual **EA Leaders Forum** (Centre for Effective Altruism, 2017c; EA Forum, 2022; Gertler, 2019); distributes enormous amounts of funding to the EA movement via **EA Funds**; creates the syllabi for university chapters' fellowships through its **Global Challenges Project** (Global Challenges Project, 2021a, 2021b, 2022a); hosts the EA community's 'community health team' which addresses conflicts in the broader community (Centre for Effective Altruism, n.d.); employs a communications specialist on behalf of the entire EA movement (Rohrig et al., 2022); and even provides coaching and media training for community members, who are cautioned against granting interviews to movement outsiders without first consulting CEA (Wise et al., 2020). Recently, CEA rebranded as **Effective Ventures (EV)** (Companies House - GOV.UK, 2022a, 2022b); now 'CEA' is a *project* of EV, along with 80K, GWWC, and various other EA initiatives including the **Atlas Fellowship**, EA Funds, Forethought Foundation, **Center for the Governance of AI**, Longview Philanthropy (formerly Effective Giving UK), **Non-Trivial Pursuits** (similar to 80K but for high-school students), and a new EA magazine called ***Asterisk*** (Effective Ventures, 2022).

2.3. From Global Poverty to 'Longtermism'?

Throughout the movement's development, EA cultivated a reputation for being primarily focused on ending global poverty by donating generously to evidence-based charities proven to save or improve the most lives per dollar. As a result, recent developments—namely, EA's embrace of 'longtermism' and seemingly new obsession with AI research and policymaking, as evidenced by Ord's latest book, *The Precipice* (Ord, 2020), Karnofsky's *Cold Takes* blog (Karnofsky, 2021b), and MacAskill's WWOTF (MacAskill, 2022e)— have caught people by surprise.

As journalists and scholars scramble to account for this 'new' version of EA—what happened to the bednets, and why are Effective Altruists (EAs) so obsessed with AI?—they inadvertently repeat an oversimplified and revisionist history of the EA movement. It goes something like this: EA was once lauded as a movement of frugal do-gooders donating all their extra money to buy anti-malarial bednets for the poor in sub-Saharan Africa; but now, a few EAs have taken their utilitarian logic to an extreme level, and focus on 'longtermism', the idea that if we wish to do the most good, our efforts ought to focus on making sure the long-term future goes well; this occurred in tandem with a dramatic influx of funding from tech scions of Silicon Valley, redirecting EA into new cause areas like the development of *safe* artificial intelligence ('AI-safety' and 'AI-alignment') and biosecurity/pandemic preparedness, couched as part of a broader mission to reduce existential risks ('x-risks') and 'global catastrophic risks' that threaten humanity's future. This view characterizes 'longtermism' as a 'recent outgrowth' (Ongweso Jr., 2022) or even breakaway 'sect' (Aleem, 2022) that does not represent *authentic* EA (see, e.g., Hossenfelder, 2022; Lenman, 2022; Pinker, 2022; Singer & Wong, 2019). EA's shift from anti-malarial bednets and deworming pills to AI-safety/x-risk is portrayed as mission-drift, given wings by funding and endorsements from Silicon Valley billionaires like Elon Musk and Sam Bankman-Fried (see, e.g., Bajekal, 2022; Fisher, 2022; Lewis-Kraus, 2022; Matthews, 2022; Visram, 2022). A crucial turning point in this evolution, the story goes, includes EAs encountering the ideas of transhumanist philosopher Nick Bostrom of Oxford University's Future of Humanity Institute (FHI), whose arguments for reducing x-risks from AI and biotechnology (Bostrom, 2002, 2003, 2013) have come to dominate EA thinking (see, e.g., Naughton, 2022; Ziatchik, 2022).

This version of events gives the impression that EA's concerns about x-risk, AI, and 'longtermism' emerged *out of* EA's rigorous approach to evaluating how to do good, and has only recently been embraced by the movement's leaders. MacAskill's publicity campaign for WWOTF certainly reinforces this perception[17]. Yet, from the formal inception of EA in 2012 (and

[17] 'It took me a long time to come around to longtermism. Over the past 12 years, I've been an advocate of effective altruism — the use of evidence and reason to help others as much as possible. In 2009, I co-founded an organization

earlier) the key figures and intellectual architects of the EA movement were intensely focused on promoting the suite of causes that now fly under the banner of 'longtermism', particularly AI-safety, x-risk/global catastrophic risk reduction, and other components of the transhumanist agenda such as human enhancement, mind uploading, space colonization, prediction and forecasting markets, and life extension biotechnologies.

To give just a few examples: Toby Ord, the co-founder of GWWC and CEA, was actively collaborating with Bostrom by 2004 (Bostrom & Ord, 2004),[18] and was a researcher at Bostrom's Future of Humanity Institute (FHI) in 2007 (Future of Humanity Institute, 2007) when he came up with the idea for GWWC; in fact, Bostrom helped create GWWC's first logo (EffectiveAltruism.org, 2016). Jason Matheny, whom Ord credits with introducing him to global public health metrics as a means for comparing charity effectiveness (Matthews, 2022), was also working to promote Bostrom's x-risk agenda (Matheny, 2006, 2009), already framing it as the most cost-effective way to save lives through donations in 2006 (User: Gaverick [Jason Gaverick Matheny], 2006). MacAskill approvingly included x-risk as a cause area when discussing his organizations on *Felificia* and *LessWrong* (Crouch [MacAskill], 2010, 2012a, 2012b, 2012c, 2012e), and x-risk and transhumanism were part of 80K's mission from the start (User: LadyMorgana, 2011). Pablo Stafforini, one of the key intellectual architects of EA 'behind-the-scenes', initially on *Felici ia* (Stafforini, 2012a, 2012b, 2012c) and later as MacAskill's research assistant at CEA for *Doing Good Better* and other projects (see organizational chart in Centre for Effective Altruism, 2017a; see the section entitled "ghostwriting" in Knutsson, 2019), was deeply involved in Bostrom's transhumanist project in the early 2000s, and founded the Argentine chapter of Bostrom's World Transhumanist Association in 2003 (Transhumanismo. org, 2003, 2004). Rob Wiblin, who was CEA's executive director from 2013-2015 prior to moving to his current role at 80K, blogged about Bostrom and Yudkowksy's x-risk/AI-safety project and other transhumanist themes starting in 2009 (Wiblin, 2009a, 2009b, 2010a, 2010b, 2010c, 2010d, 2012). In 2007, Carl Shulman (one of the most influential thought-leaders of EA, who oversees a $5,000,000 discretionary fund at CEA) articulated an agenda that is virtually identi-cal to EA's 'longtermist' agenda today in a *Felici ia* post (Shulman, 2007). Nick Beckstead, who co-founded and led the first US chapter of GWWC in 2010, was also simultaneously engaging with Bostrom's x-risk concept (Beckstead, 2010). By 2011, Beckstead's PhD work was centered on Bostrom's x-risk project: he entered an extract from the work-in-progress, entitled "Global Priority Setting and Existential Risk: Crucial Ethical Considerations" (Beckstead, 2011b) to FHI's "Crucial Considerations" writing contest (Future of Humanity Institute, 2011), where it was the winning submission (Future of Humanity institute, 2012). His final dissertation, entitled *On the Overwhelming Importance of Shaping the Far Future* (Beckstead, 2013) is now treated as a foundational 'longtermist' text by EAs.

Throughout this period, however, EA was presented to the general public as an effort to end global poverty through effective giving, inspired by Peter Singer. Even as Beckstead was busy writing about x-risk and the long-term future in his own work, in the media he presented himself as focused on ending global poverty by donating to charities serving the distant poor (Beckstead & Lee, 2011; Chapman, 2011; MSNBC, 2010). MacAskill, too, presented himself as dog-gedly committed to ending global poverty and quickly became the public-face of EA through his appearance on "Intelligence²" (Intelligence Squared, 2015), a TEDx talk entitled "Want to Make a Difference? Don't Work for a Charity" (MacAskill, 2015e), and his regular contributions to *Quartz* between 2013-2015 (MacAskill & MacAskill, 2015; MacAskill, 2013a, 2013b, 2013d).

that has raised hundreds of millions of dollars to help pay for bed nets to protect families against malaria and medicine to cure children of intestinal worms, among other causes. These activities had a tangible impact. By contrast, the thought of trying to improve the lives of unknown future people initially left me cold' (MacAskill, 2022a); 'I didn't always buy these ideas. For much of my life, I spent my energy on more visceral problems - my first book, Doing Good Better, was all about how we can most effectively improve the lives of the extreme poor, such as by funding anti-malarial bednets. But ultimately the arguments for giving moral concern to future generations and for working to make their lives go better won me over. Those arguments are what What We Owe The Future is about' (MacAskill, 2022d); 'Drawing on what I have learned, I have tried to write the case for longtermism that would have convinced me a decade ago' (MacAskill, 2022e, p. 7).

[18] In the acknowledgments of *The Precipice*, Ord writes 'the greatest influence on me at Oxford has been Nick Bostrom [...]. I think he introduced me to existential risk the day we met, just after we both arrived here in 2003; we've been talking about it ever since' (Ord, 2020, p. 220).

2013e, 2014a, 2014b, 2014d, 2015a, 2015c, 2015d). In 2015, MacAskill published what has become the most popular introductory text to EA, *Doing Good Better* (MacAskill, 2015b), launching his career and making him the most visible figurehead of the EA movement. Meanwhile, Peter Singer—who apparently believed that his new heirs' interest in AI-safety/x-risk was merely a thought experiment, not a cause they actually endorsed (Beckstead et al., 2013)[19]—continued his full-throated support for the EA movement, granting credibility and spurring EA's reputation as resolutely dedicated to ending global poverty through his book *The Most Good You Can Do: How Effective Altruism is Changing Ideas about Living Ethically* (Singer, 2015), his 2013 TED talk, "The Why and How of Effective Altruism" (Singer, 2013), and his open online course about EA (Singer, n.d.).

EAs *themselves* refer to this as *public-facing* EA — the global-poverty-focused version of EA aiming to build support and credibility amongst the general public and new recruits (Centre for Effective Altruism, 2017d, p. 15). Since the mainstream understanding of EA comes almost solely from public-facing content, observers have treated EA as a controversial but nonetheless earnest approach to ending global poverty (along with a less-prominent focus on animal welfare). Reflecting this selective reading, what little academic literature exists on EA treats it as a synonym for Singer's effective giving concept: scholars have raised issues with EA's narrow methodological preference for quantitative metrics and its tendency to treat randomized controlled trials (RCTs) as the 'gold standard' of aid effectiveness data (Brown, 2016; Cochrane & Thornton, 2016; Deaton, 2015; Gabriel, 2017; Muyskens, 2017); debated EA's apparent lack of engagement with historical context, politics, and the 'systemic change' issue, i.e., that EA treats *symptoms* but leaves in place—or even buttresses—the structures that produce inequality (Berkey, 2018; Clough, 2015; Crary, 2021; Eikenberry & Mirabella, 2018; Kissel, 2017; Lechterman, 2020; Rubenstein, 2016; Srinivasan, 2015; Syme, 2019); and argued that EA ignores important classes of interventions because they are not amenable to measurement (Côté & Steuwer, 2023; Read, 2018). From the perspective of moral philosophy and applied ethics, scholars have analyzed EA's principles regarding moral obligations to assist others (Berg, 2018; Igneski, 2016; Isaacs, 2016; Krishna, 2016; Law et al., 2021; Lim, 2019; Mihailov, 2022; Pellegrino, 2017; Pummer, 2016; Schaller, 2018; Skelton, 2016). The present article does not engage with these arguments — not because they are without merit, but because they do not address the *actually-existing* EA movement.[20]

3. Crafting Strategic Ambiguity: Tactical Polyvalence, Glittering Generalities and Frame Appropriation

How does EA continue to be focused on ending global poverty to some people, and about AI-safety/x-risk and safe-guarding the long-term future, to others? The answer lies in the structure of EA rhetoric: EA has multiple discourses running simultaneously, using the same terminology to mean different things depending on the target audience. The most important aspect of this double rhetoric, however, is not that it maintains two distinct arenas of understanding, but that it also serves as a credibility bridge *between them*, across which movement recruits (and, increasingly, the general public) are led in incremental steps from the less controversial position to the far more radical position.

[19] Singer's *The Most Good You Can Do* (2015) contains a final chapter that goes quite deeply into Bostrom's x-risk argument and MIRI's case for ensuring the development of friendly AI, explicitly referencing Bostrom and MIRI executive director Luke Muehlhauser (who joined OpenPhil shortly after); however, Singer presents these ideas as merely 'the further reaches of conversations in which philosophers and some of the more philosophically minded effective altruists engage', adding that 'If these discussions lead in strange directions, never mind' (Singer, 2015, p. 177). Since 2019, Singer has registered increasing alarm at EA's embrace of 'longtermism'/x-risk (see, e.g., Singer, 2021; Singer & Wong, 2019 from 00:05:00 to 00:05:56).

[20] Notable exceptions where authors have noted links between EA, the tech sector, and transhumanism are Srinivasan (2015), Impett (2018), and Pinto (2019). In recent years, a former EA/'longtermism' adherent turned critic has also highlighted EA's origins amongst the transhumanists (Torres, 2020), and in 2021 two scholars published an excellent critique of the x-risk concept, expressly linking it to EA (Schuster & Woods, 2021). However, such arguments failed to permeate mainstream scholarship on EA until WWOTF's launch in 2022.

3.1. Concepts: Tactical Polyvalence, Frame Appropriation, and Glittering Generalities

A starting point for understanding EA's multiple discourses is the Foucauldian concept of ***tactical polyvalence***. Foucault explains tactical polyvalence as 'the shifts and reutilizations of identical formulas for contrary objectives' (Foucault, 1978, p. 100), such that discourses become 'tactical elements' wherein 'there can exist different and even contradictory discourses within the same strategy; they can, on the contrary, circulate without changing their form from one strategy to another, opposing strategy' (Foucault, 1978, pp. 101-102). For example, in her work on human rights and neoliberalism, Jessica Whyte uses the concept of tactical polyvalence to explore how the language of human rights has been easily co-opted to defend power, wealth, and acts of war (Whyte, 2014, 2019): claims to be *protecting women and children* may refer to efforts that help women and children fleeing violence, but may just as easily refer to the very acts of military intervention that cause the violence from which women and children must flee (Whyte, 2014).

Frame appropriation is one way that a specific term may come to be tactically polyvalent. Frames articulate views of reality in ways intended to persuade and mobilize participants, by conveying partial or selected meanings in support of a particular action or agenda (Snow, 2004). Frame appropriation occurs when actors transform an existing discursive frame to suit their needs: 'rather than inventing a new and competing frame' actors redefine an existing frame 'from within' (Joachim & Schneiker, 2012, p. 368).

Another way to employ tactical polyvalence is through ambiguous words with positive connotations. In their classic study of propaganda, Lee and Lee (1939) refer to such linguistic devices as ***glittering generalities***, words met with approval by audiences, but without examining the speaker's meaning closely. This differs from frame appropriation—which seeks to invert or transform an existing frame that has been used for a specific, albeit contested and unstable, meaning—in that the words used are free-floating and ambiguous: a word like 'good' seems positive, but what a speaker thinks is 'good' is subjective. Speakers who employ glittering generalities leave precise meanings unspoken ('doing good', 'helping others', 'tackling the world's most pressing problems'); listeners must fill in the blanks (what constitutes 'good'? What kind of 'help'? What is 'most pressing'?) based on *what they assume* the speaker means. Listeners extrapolate from examples the speaker used or based on *their own* beliefs.

The subtle distinction I draw between frame appropriation and glittering generalities boils down to this: frame appropriation uses the polyvalence of words to say one thing while meaning a *specific other* thing; glittering generalities leave meanings open and vague so that there is always plausible deniability about what the speaker *really* intended. In practice, however, frame appropriation and glittering generalities intermingle and the distinction is less clear-cut, since the ambiguity afforded by glittering generalities supports the creation of specific frames that can then be easily appropriated vis-à-vis the glittering generality.

In this section, I focus on glittering generalities in EA discourse, although this is not the only form of tactical polyvalence employed by the movement. What matters is that the ambiguity of phrases underpinning EA rationales is strategic, facilitating multiple interpretations while also supplying recourse to plausible deniability to reconcile gaps between what one *said* and what one *meant*. Most importantly, the repeated use of the same words and concepts serves as a *credibility bridge*, starting from relatively non-controversial claims and then building toward increasingly controversial ones.

3.2. EA's Glittering Generalities and the Public-Facing/Core Divide

CEA defines *Effective Altruism* as the project of 'using evidence and reason to figure out how to benefit others as much as possible, and taking action on that basis,' and *the EA community* as 'a global community of people who care deeply about the world, make benefitting others a significant part of their lives, and use evidence and reason to figure out how best to do so' (Centre for Effective Altruism, 2017b). MacAskill explains that these definitions are *non-normative* (not prescribing a specific action), *maximizing* (more good is better than less good, e.g., if we can save 10 or 100 people with the same resources, we ought to save 100), *science-*

aligned (rigorously evidence-based or arrived at via careful analysis), and broadly *welfarist and impartial* (aiming to do good for others based on the cosmopolitan principle that all persons have equal moral value, regardless of race, nationality, culture, religion, gender, etc.) (MacAskill, 2019). This definition accords with the idea that EA is a question ('How can I do the most good, with the resources available to me?' (Toner, 2014)), or method ('a particular way of trying to solve problems' (THINK, 2013, p. 2)), *not* an answer or specific cause. Crucially, this definition is constructed out of glittering generalities: what one thinks is 'effective', what one considers 'altruistic', and what it means to 'benefit others as much as possible' remain open to subjective interpretation.

This definition of EA (and the foundational tenets associated with it, such as *cause-neutrality*, *impartiality*, and *openness to unusual ideas*) creates a discursive reservoir of plausible deniability, which can be dipped into as needed: EA is presented as an effort to end global poverty when it is *expedient* to do so (in the mainstream media, when talking to left-wing/progressive academics, when presenting EA ideas to a new audience or potential recruits), while in a different context (communications with core EAs, or, as is happening currently, to explain the sudden shift to 'longtermism') the initial focus on global poverty is explained as merely *an example* used to illustrate the concepts — not the *actual* cause endorsed by most EAs (see e.g., Bankman-Fried, 2012; Hutchinson, 2016; Todd, 2017). This is where the plausible deniability afforded by glittering generalities comes into play: all that gnashing of teeth about the distant poor in sub-Saharan Africa—the children who died because *you* bought a new laptop or a cinema ticket or donated to the animal shelter, instead of donating that money to GiveWell-recommended charities[21]—*that was just an example to illustrate the principle*. As Jonas Vollmer[22] candidly explained when asked about his EA organization Raising for Effective Giving, 'REG prioritizes long-term future causes, it's just much easier to fundraise for poverty charities' (Vollmer, 2018).

The firebrand rhetoric EAs employed in public-facing EA concretized the public's opinion of EA as a group of idealistic students, perhaps misguided in their approach to ending poverty through RCTs and private charity, but ultimately a well-intentioned force for 'good' in the world (Brest, 2015). The public understanding was that when EAs talked about 'doing good', they meant enacting a transfer of resources from people in wealthy countries to the global poor, and 'evidence and reason' meant RCTs. Once the public's trust in EA as a science-based effort to eliminate global poverty was established, EAs could dip into their discursive reservoir, relying on glittering generalities to communicate on two different registers: one where 'the most pressing problems' were understood to refer to poverty and disease, and another where the 'most pressing problems' were understood as extinction from unaligned AI/biotechnology or the threat of technological stagnation that would lead to humans never creating superintelligent AI, never eliminating suffering through genetic engineering, never colonizing space. If accused of hypocrisy or of misleading the public about their true intentions, they could always fall back on glittering generalities: they never claimed to be working to end global poverty, they only claimed to be doing good, helping others, pursuing the most impact.

Impartiality in public-facing EA is derived from Peter Singer's drowning child thought experiment[23] and argues that *physical* (*geographical*) *distance* does not absolve us of the duty

[21] EA's public-facing discourse often expresses indignation about wasteful spending: 'If you spend eight hundred dollars on a laptop, that's one African kid who died because you didn't give it to charity' (Alexander, 2012); 'Money that you won't even miss could be saving lives right now if you put it to that purpose instead of, say, home improvement or collecting action figures' (Elmore, 2016); 'If I spend my money going to see a movie, does that help people in poor countries as much as if I provide them with bednets?' (Ord, in MacFarquhar, 2015, p. 95); 'choosing whether you buy an extra burrito versus donating a dollar to charity can also make a big difference in how much you can help others' (Hurford, 2012).
[22] Vollmer has been involved with EA since 2012, when he helped co-found/lead various interrelated organizations in the Swiss and German EA scene, including GBS-Switzerland, EA Switzerland/EA Geneva, the EA Foundation (now renamed the Center on Long-Term Risk, CLR), and Raising for Effective Giving (REG). Vollmer has also held positions of considerable influence in terms of allocating funding, as a fund manager for the EA Foundation/CLR Fund, the Center for Emerging Risk Research (now renamed Polaris Ventures), and CEA's EA Funds. Currently Vollmer is the executive director of CEA's Atlas Fellowship program.
[23] Singer's thought experiment illustrates the moral obligation of people in wealthy countries to donate to organizations that help the very-poor overseas: given the choice of wading into a shallow pond to save a drowning child at no risk to our own well-being (aside from our shoes getting ruined), virtually everyone chooses to save the child, for

to assist: all people have moral worth, regardless of *where* they live. If you truly care about helping people impartially, you should prioritize the 'distant poor' in sub-Saharan Africa where you can help thousands of people for the cost of helping just one person in a wealthy country. Impartiality in core EA, however, refers primarily to *temporal distance*: all people have equal moral worth, regardless of *when* they live, including (especially) future generations that will come into existence millions (or even billions) of years from now. In fact, (this view says) we ought to care *more* about future generations since not only are they vastly larger in aggregate, but they are also utterly disenfranchised by our current institutions — they have no voice, no vote, no way to make their needs known (80000 Hours, 2017; Delo, in Jacobs, 2019; MacAskill, in Levitz, 2022; Longview Philanthropy [formerly Effective Giving UK], 2020; MacAskill, in Wiblin et al., 2020). To be truly impartial about helping others, one should adopt the 'longtermist' perspective, prioritize the long-term future, and care most about affecting the welfare of future generations simply because it helps many more individuals (Moorhouse, 2022).

Cause-neutrality in public-facing EA means that if you want to save lives, you should not donate to causes based on familiarity or prior personal involvement — for instance, the local animal shelter or to research on a rare disease that claimed the life of a loved-one; it means objectively seeking out where your resources can do the most good, using expected value calculations, and recognizing that although leukemia and breast cancer sound scarier than diarrhea and malaria, in fact, you likely save more lives by donating to the Against Malaria Foundation (AMF) than to cancer research (Boudry, 2017; Deere, 2016). But when cause-neutrality is invoked in core EA, it means that commonly accepted causes like reducing poverty or improving health in developing countries are probably not the best way to help others the most. Expected value calculations show that that high-risk/high-reward scenarios, like x-risk reduction and AI-safety, have exceptionally high expected value and therefore should be prioritized.

Directly related to cause-neutrality is the EA injunction to remain **open to taking unusual ideas seriously**. Taking ideas seriously features prominently in EA's off-the-shelf introductory materials for student groups, where it appears to be little more than a synonym for critical thinking. EAs are urged to question whether and under what conditions one can reliably trust one's intuitions, to question the status quo, and to take ideas seriously rather than simply dismissing unusual ideas — *particularly* when they seem strange or unintuitive (Global Challenges Project, 2022b). This accords with many values that are prized by academics and intellectuals, such as questioning the status quo; being self-reflexive about one's own positionality, biases, and the epistemic community in which one is enmeshed; and recognizing that many ideas that seem commonplace today started out sounding unusual or counterintuitive – these are all hallmarks of critical thinking. Applied in public-facing EA, openness to unusual or unintuitive ideas means understanding that charities working on unsexy causes like diarrhea and trachoma operations are actually doing more *good* (in EA's cost-effectiveness terms) than interventions that are sensationalized in the media, like the ALS "Ice-Bucket Challenge" or mainstream charities like animal shelters or the Make-A-Wish Foundation. It may seem strange to think this way, but it is essential if we are serious about helping others the most (MacAskill, 2017).

In core EA, taking unusual ideas seriously still means rejecting one's intuitions and gut feelings, but now extended *beyond* the reach of evidence or checks against reality. The frequent invocation in EA to take ideas seriously and to distrust one's instincts ends up yielding the opposite of open-mindedness, closing EAs' minds to insights that do not support EA conclusions and to external criticism of the EA worldview. Specifically, it means taking 'longtermism' seriously. Rather than encouraging critical thinking, in core EA the injunction to take unusual ideas seriously means taking *one very specific set* of unusual ideas seriously, and then providing increasingly convoluted philosophical justifications for why those particular ideas matter most.

a child's life is worth more than a pair of shoes (Singer, 1972, 1997). Singer asks why we do not recognize the same duty toward those who are physically far away—the 'distant poor'— dying of famine or illness in poor countries where we can also take action that poses no risk to our own well-being, in the form of modest monetary donations (equivalent, say, to the price of a new pair of shoes), thereby saving the lives of many children. If we agree that all people have equal moral worth, he argues, geographical distance does not absolve us of the obligation to assist since we can rely on aid organizations to overcome the distance for us: 'If we accept any principle of impartiality, universalizability, equality, or whatever, we cannot discriminate against someone merely because he is far away from us' (Singer, 1972, p. 232)

This is framed as an exercise in 'open truthseeking' — something that sounds entirely unobjectionable to those who value reflexivity and critical thinking (Effectivealtruism.org, 2022). But once one understands the double-rhetoric at play, the injunction for remaining open to unusual ideas and being willing to radically change one's views actually just means being open to taking x-risk/AI-safety/'longtermism' seriously. It may be *uncomfortable* to prioritize the creation of artificial general intelligence (AGI; also known as smarter-than-human AI or 'superintelligence') so as to ensure (post)humanity colonizes the galaxy, rather than prioritizing a familiar cause that helps people who are presently suffering. It may be *unintuitive* to care more about the potential suffering of hypothetical future people than about those who are hungry, or poor, or suffering from easily treated diseases today. But if (EAs argue) we *truly* care about doing the most good impartially and cause-neutrally, we need to take seriously the idea that the expected value of even the tiniest reduction of existential risk or positive trajectory change affecting the long-term future far outweighs any other cause. It requires considerable bullet-biting and the ability to go against one's deepest instincts about how best to help others, but this is what makes EAs special:

> I feel like a big part of the edge of the EA and rationality community is that we follow arguments to their conclusions even when it's weird, or it feels difficult, or we're not completely sure. We make tradeoffs even when it feels really hard - like working on reducing existential risk instead of helping people in extreme poverty or animals in factory farms today. (Holness-Tofts, 2021)

Thought-leaders help with the transition by acknowledging how disorienting it can be to accept the 'longtermist' doctrine. GiveWell co-founder Holden Karnofsky admits that his belief that we are living in the most important century (wherein we either develop superintelligent AI, digital minds, and put humanity on the right trajectory for a flourishing, stable intergalactic civilization, or enter a period of stagnation, decline and finally extinction) has

> a wacky, sci-fi feel. It's very far from where I expected to end up when I set out to do as much good as possible. But part of the mindset I've developed through GiveWell and Open Philanthropy is being open to strange possibilities, while critically examining them with as much rigor as possible. (Karnofsky, 2021a)

In public-facing/grassroots EA, **'doing good'** is typically understood to mean helping the needy and vulnerable. Doing good means effective giving, i.e., donating generously, and making sure that your donation goes to the most effective charities in terms of lives saved or improved per dollar. Even the poor in wealthy countries who earn minimum wage are still among the wealthiest people in the world and ought to do their part (West, in Zhang, 2016). Donating to a cause in a wealthy country—like scholarships for students at elite university programs in the US or UK[24], the arts or local animal shelter[25], or environmentalism[26]—is not just wasting a donation but literally condemning people in poor countries to death. In core EA, doing good means working to mainstream core EA ideas (i.e., promoting 'longtermism', AI-safety, x-risk/global catastrophic risk reduction in academia, in the media, in the public consciousness, etc.) and furthering the interests of the tech sector by working on technical AI or biotech research, or on policy entrepreneurship for emerging technologies.

Directly related to this subjective and variable understanding of what constitutes 'doing good' is observable divergence in what/who is considered a legitimate beneficiary to receive EA donations. When it comes to the **target recipients** of donations in public-facing/grassroots EA, people who consider themselves 'effective altruists' are *donors to* EA causes. These individuals often take the GWWC pledge, and commit to donate 10% of their income to effective charities — typically understood to be GiveWell's top recommendations, with the Against Malaria Foundation (AMF), Schistosomiasis Control Initiative (SCI), Deworm the World, and GiveDirectly

[24] Criticizing a large donation to Stanford University supporting the Knight-Hennessy scholarship for graduate students, EA reporter Dylan Matthews stated that Stanford ought to explain 'why helping 100 elite students afford a few years of graduate school is better for the world than saving thousands of lives or helping tens of thousands of people escape poverty' (Matthews, 2016).
[25] 'Over a lifetime of giving, the choice to donate to one cause rather than another […] is the decision to let tens or hundreds of people die, in exchange for, say, one additional opera performance or saving several stray dogs' (MacAskill, 2013c)
[26] 'Donating $2000 to a charity planting trees, means an extra person dies of malaria. Brushing the trade-offs aside is not an option' (Todd, 2013b)

being the most popular (and the most-routinely-cited in public-facing EA content). Here, the beneficiaries of EA donations are disadvantaged people in the poorest countries of the world. In stark contrast to this, the target recipients of donations in core EA are the EAs *themselves*. Philanthropic donations that support privileged students at elite universities in the US and UK are suddenly no longer one of the *worst* forms of charity but one of the best. Rather than living frugally (giving up a vacation, skipping a trip to the cinema or a restaurant, getting rid of an expensive gym membership) so as to have more money to donate to AMF, providing such perks is now understood as *essential* for the well-being and productivity of the EAs, since they are working to protect the entire future of humanity (Jeyapragasan, 2021).

Table 1: The Double-meanings of EA Rhetoric: Turning glittering generalities into scripts for different audiences

	Public-facing /grassroots EA	Core EA
Cause neutrality, being open to unusual /counterintuitive ideas	Reject intuitions, calculate expected value What initially 'sounds' like a 'good' way to help others probably isn't the 'best' way: - To improve school outcomes, deworming pills > books - To help children, AMF > Make-a-Wish - To help blind people, trachoma operations > Guide dogs - Work to end factory farming > donating to animal shelter - Work on Wall St. and earn to give > work directly for a non-profit Support efforts that may be unintuitive but have unusually high expected value (work on ending diarrhea, malaria; provide cash grants for the poor; work toward ending factory farming; become veg*n)	Reject intuitions, calculate expected value What initially 'sounds' like a 'good' way to help others probably isn't the 'best' way: - Reducing x-risk to the long-term future of humanity > short term improvements in global health/ poverty reduction - Relieving wild animal suffering > ending factory farming - Helping elite students get advanced degrees > helping the global poor - AI-safety/x-risk > GiveWell charities Support efforts that may be unintuitive but have unusually high expected value (help EAs earn machine learning/ technical AI degrees; fund personal assistants for impactful community-members; ensure EAs wellbeing)
Doing good, helping others	Helping the poor, needy, and vulnerable by supporting organizations that directly provide assistance	Helping future populations by supporting the members of the EA community in their effort to save the world
Tackle the world's most pressing problems	End global poverty and inequality Prevent children from dying from diseases that are easily/cheaply preventable or curable End animal suffering on factory farms	Prevent unsafe/unaligned AGI; promote beneficial AGI Prevent unsafe biotechnology; promote beneficial biotechnology Prevent technological stagnation; ensure space colonization to preserve humanity Spread EA values
Impartiality **(caring about others equally)**	All people matter, regardless of *where* they were born. The *geographically distant* ('distant poor') who live in poor countries overseas are the most vulnerable population and should be prioritized. All sentient beings matter, including non-human animals	All people matter, regardless of *when* they are born. The *temporally distant* who are yet-to-be-born and who constitute the aggregate future of humanity are the most vulnerable population and should be prioritized. All sentient beings matter, including non-human animals/ insects and artificial sentience such as digital minds
Evidence & Reason	*Evidence-based:* RCTs, quantitative metrics, evidence of impact produced through scientifically-sound evaluations	*Reasoning-based:* Expected-value calculations and Bayesian extrapolations from subjectively defined priors
The most effective charities	GiveWell-recommended aid interventions (e.g. AMF, GiveDirectly, SCI, Deworm the World, Helen Keller International)	AI-safety/x-risk reduction organizations (MIRI, FHI, Berkeley CHAI, OpenAI, Redwood Research, Ought) Think-tanks working on AI/tech policy (GovAI, CSET, Johns Hopkins CHS, Nuclear Threat Initiative) EA Community-building Supporting individual EAs to develop high-impact careers

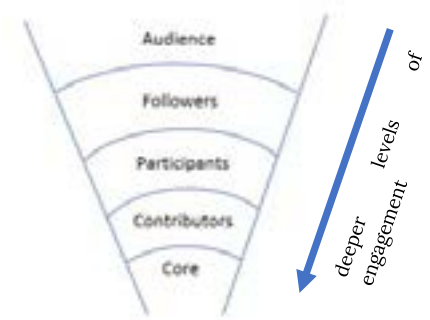
4. The Funnel Model

The two versions of EA described in the previous section result from a deliberate strategy, which can be understood as serving two aims: (1) it enables EA to maintain a (relatively) non-controversial public face, and (2) provides a social technology for incrementally introducing (or converting) newcomers to core EA ideas.

CEA explains the different levels of engagement as a series of concentric circles: outside the largest circle is 'audience', the first ring is 'followers', then 'participants', 'contributors', 'core', and, at the very center, 'leadership' (Centre for Effective Altruism, 2019a). The *audience* is members of the general public interested in mainstream EA ideas; *followers* have had (limited) engagement with EA and understand the basic principles; *participants* are more engaged than followers, use EA research to inform their giving decisions and often have taken the GWWC pledge; *contributors* have typically taken the GWWC pledge but also contribute to the growth of the EA movement through in-depth engagement with the EA community (e.g., attend EA Global, do an internships at EA/EA-aligned organizations, post on the EA *Forum*, run local EA groups); the *core* has 'expert-level' understanding of key EA ideas and devote their resources to acting on these ideas, particularly in terms of choosing EA-recommended careers (e.g., direct work on EA community-building or in an EA research organizations) rather than by donating; and finally, *leadership* figures, who demonstrate 'an understanding and devotion similar to the core, but they are also leaders of major effective altruist organizations, or are leaders of the intellectual development of the community' (Centre for Effective Altruism, 2019a).

For public relations, outreach, recruitment, and community-building, CEA imagines the circles of engagement as rungs in a funnel (borrowing from a marketing strategy). In EA's '**funnel model**' (author's rendering, Figure 1, below), the goal is to move as many people as possible from the top of the funnel to the core (Centre for Effective Altruism, 2018b).

Figure 1 CEA's 'Funnel Model', based on Centre for Effective Altruism (2018b)



When someone first hears about EA, they enter the top of the funnel; as they learn more, they may go to deeper, stay at the same level, or decide that they are no longer interested and leave ('leakage'). The funnel is wide at the top, representing ideas that potentially attract many people who have casual knowledge but limited/low engagement with the movement (audience to participants), i.e., people who are interested in 'grassroots effective altruism' based on public-facing materials and popular media (e.g. *Doing Good Better*), which stress Singer's concept of effective giving, global poverty, and animal welfare (note: these causes are now referred to in the community as 'neartermist' causes to distinguish them from 'longtermist' causes). Many people who learn about EA do not go deeper than this, which explains the widespread but mistaken idea that EA just refers to the idea of evaluating global aid/development interventions' cost-effectiveness or donating 10% of one's income to GiveWell-recommended global poverty charities. The bottom of the funnel (contributors to core) is smaller, but consists of people highly committed to EA's more esoteric ideas ('longtermism', AI-safety/x-risk, prediction and forecasting markets, wild animal/insect suffering etc.), and to the EA community. Core

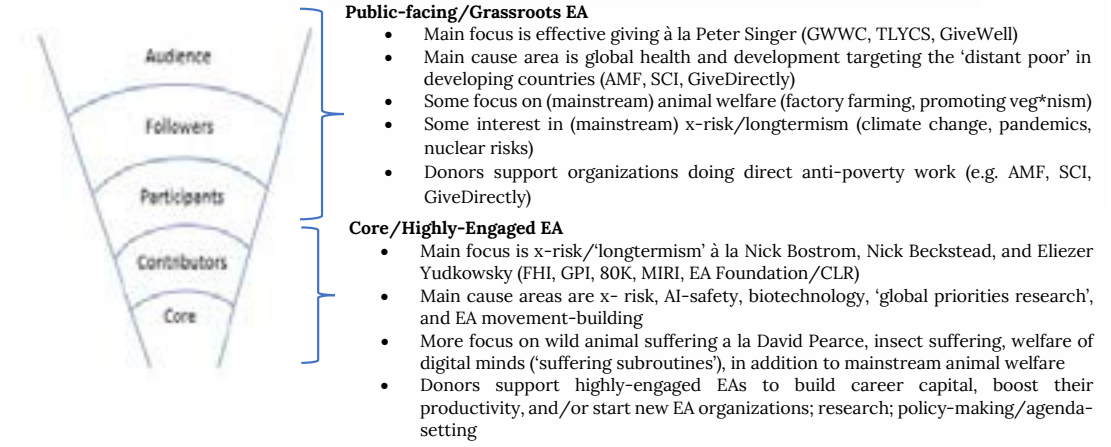
community members, also referred to as 'highly-engaged EAs' or 'HEAs' (Centre for Effective Altruism, 2021b, p. 5; Dalton, 2021), typically seek direct work in EA and EA-aligned organizations, where their 'path to impact' is via their careers rather than through donations.

At first, CEA concentrated outreach on the top of the funnel, through extensive popular media coverage, including MacAskill's *Quartz* column and book, *Doing Good Better*, Singer's TED talk, and Singer's *The Most Good You Can Do*. The idea was to create a broad base of poverty-focused, grassroots effective altruists to help maintain momentum and legitimacy, and act as an initial entry point to the funnel, from which members sympathetic to core aims could be recruited. The 2017 edition of the movement's annual survey of participants (conducted by one of the movement's major organizations, Rethink Charity) notes that this is a common trajectory:

> new EAs are typically attracted to poverty relief as a top cause initially, but subsequently branch out after exploring other EA cause areas. [...] In other words, the top of the EA outreach funnel is most relatable to newcomers (poverty), while cause areas toward the bottom of the funnel (AI) seem more appealing with time and further exposure. (Hurford & Barnett, 2017)

EA's glittering generalities form the glue—the credibility bridge—that holds EA's different rungs together; they are the rails along which new recruits ride from the top of the funnel to the core. At first, doing good means taking the GWWC pledge and donating to GiveWell/ACE-recommended charities; a bit deeper in the funnel, doing good means helping build the EA community, becoming vegan or worrying about x-risks like climate change and pandemics; at the core, doing good means accepting the overwhelming importance of AI-safety/x-risk/'longtermism', and working full-time on things like promoting EA/'longtermism' at universities, becoming a personal assistant (PA) to a high impact EA, studying machine learning to work on technical AI-alignment, or pursuing a career in emerging technology policy. (see figure 2, below)

Figure 2: Differential Portrayal of EA in the top versus bottom of the funnel



EAs equate moving through the funnel with Yudkowsky's concept of crossing 'inferential distances':

> A clear argument has to lay out an inferential pathway, starting from what the audience *already knows or accepts*. If you don't recurse far enough, you're just talking to yourself. If at any point you make a statement without obvious justification in arguments you've previously supported, the audience just thinks you're crazy. (Yudkowsky, 2007)

While Yudkowsky's formulation offers a concise heuristic for understanding how new knowledge claims must build upon previously-accepted knowledge to be credible, it can also function as a template for persuading listeners to accept increasingly outlandish propositions without stepping back to evaluate them against reality. As long as each new claim can be shown to *logically* proceed from a previously-accepted claim, it is possible to start in the realm of the banal ("I think it's good to help others") and wind up in fairly extreme territory ("I should become an investment banker so that I can pay for students doing AI research to have personal assistants, because this will help ensure humanity colonizes the galaxy"). This is not merely a bridging of inferential gaps, but *manufacturing a credibility bridge* – a process of artificially generating *the appearance* of links between a series of claims, held together by shared vocabulary and a set of principles that remain consistent at every point on the bridge (EA's glittering generalities). Those who proceed along this credibility bridge and go deeper into the EA funnel are not gaining new knowledge so much as being indoctrinated into a belief system.

Many community-members explain that they became convinced about the importance of AI-safety/x-risk after engaging with influential thinkers and ideas in the community, thus the strategy appears effective:

> Initially I cared a lot about global poverty, but by engaging with the effective altruism community I began to see the overwhelming importance of animal well-being and far future issues such as existential risks. (Mindermann, 2013)

> I came to EA for the GiveWell top charities [...] It took me several years to come around on the longtermism/x-risk stuff, but I never felt duped or bait-and-switched. Cause neutrality is a super important part of EA to me and I think that naturally leads to exploring the weirder/more unconventional ideas. (User: Jeremy, 2022)

> My experience, and the experience of many others, is that we joined EA because it seemed like the best way to do good in the neartermist cause areas we were already interested in, learned about longtermism through our involvement with EA, and eventually shifted our focus to include longtermism. Essentially, neartermist causes served as an on-ramp to EA (and to longtermism). (Garth, 2022)

> On one hand, basically all the smart EA people I trust seem to be into longtermism; it seems well-argued and I feel a vague obligation to join in too. On the other, the argument for near-term evidence-based interventions like AMF is what got me [...] into EA in the first place. (User: akrolsmir, 2020)

Yet others question the efficacy of this strategy, seeing the global poverty entry-point as unnecessary and potentially dishonest:

> I became an EA in 2016, and it the time, while a lot of the "outward-facing" materials were about global poverty etc, with notes about AI safety or far future at much less prominent places. I wanted to discover what is the actual cutting edge thought, wen[t] to EAGx Oxford and my impression was the core people from the movement mostly thought far future is the most promising area, and xrisk/AI safety interventions are top priority. I was quite happy with that, the reasoning why focus on far future seems sound to me. [...] However, I was somewhat at unease that there was this discrepancy between a lot of outward facing content and what the core actually thinks. With some exaggeration, it felt like the communication structure is somewhat resembling a conspiracy or a church, where the outward-facing ideas are easily digestible, like anti-malaria nets, but as you get deeper, you discover very different ideas...so at the end you are not that likely to think donating to AMF is the best thing to do. (Kulveit, 2018)[27]

[27] Jan Kulveit leads the European Summer Program on Rationality (ESPR), an all-expenses paid EA program that targets mathematically-gifted teenagers and serves as an entry point to EA's AI-safety talent pipeline (European Summer Program in Rationality, 2022). ESPR was originally created as the European wing of CFAR's Summer Program on Applied Rationality and Cognition (SPARC), but was spun off as a separate organization in 2019. While a project of CFAR, ESPR received $340,000 from OpenPhil (Open Philanthropy, 2017), and since becoming an independent organization, ESPR has received $5,225,000 from OpenPhil (Open Philanthropy, 2019, 2022k). Kulveit also co-leads another EA summer program focused on AI-safety called the Human-Aligned AI Summer School. These programs are funded through OpenPhil's 'longtermist EA' grantmaking and CEA's EA Funds.

Prominent EA community-member and blogger Ozy Brennan has argued that EA would be better off if everyone was honest about their preferred cause area:

> [W]hen you are speaking as an effective altruist– don't get complicated, don't get clever. Just say what you think the best cause area or charity or career is. Every time you think to yourself "well, I think AI risk is more important, but it'll turn people off, so I should probably say the Against Malaria Foundation," the effective altruism movement takes one more step towards being the same as any other group of charitably-minded nerds. […] A lot of introductory effective altruism material uses global poverty examples, even articles which were written by people I know perfectly fucking well only donate to MIRI. (Brennan, 2017)

While Yudkowsky—who once admitted 'I regard the non-x-risk parts of EA as being important only insofar as they raise visibility and eventually get more people involved in, as I would put it, the actual plot' (Yudkowsky, 2013)—has argued the opposite:

> As I've said repeatedly, xrisk cannot be the public face of EA, OPP [OpenPhil] can't be the public face of EA, animal altruism can't be the public face of EA. Only 'sending money to Africa' is immediately comprehensible as Good and only an immediately comprehensible Good can make up for the terrible PR profile of maximization or cause neutrality. And putting AI in there is just shooting yourself in the foot. (Yudkowsky, 2015)

Scott Alexander, the author of the influential rationalist blog *Slate Star Codex* and a much-revered figure in the EA community—and author of two influential public-facing EA texts, one where he accuses donors to an art museum of letting thousands of people in poor countries die in exchange for a painting (Alexander, 2013 [2010]), and another where he states that anyone who spends $800 on a new laptop is choosing to let one African child die (Alexander, 2012)—has acknowledged that he actually supports the AI-safety/x-risk cause area, but believes EAs should not mention it in public-facing material:

> Existential risk isn't the most useful public face for effective altruism – everyone inc[l]uding Eliezer Yudkowsky agrees about that (Alexander, 2015b)

> Several people have recently argued that the effective altruist movement should distance itself from AI risk and other far-future causes lest it make them seem weird and turn off potential recruits. Even proponents of AI risk charities like myself agree that we should be kind of quiet about it in public-facing spaces. (Alexander, 2015a)

Rob Bensinger (the research communications manager at MIRI and prominent EA movement member) argued in 2016 for a middle approach, acknowledging that AI was not the optimal public-face for EA, but cautioning against EA becoming locked into the default public-facing stance wherein it is treated synonymously with GiveWell-recommended charities:

> [I]n fairness to the "MIRI is bad PR for EA" perspective, I've seen MIRI's cofounder (Eliezer Yudkowsky) make the argument himself that things like malaria nets should be the public face of EA, not AI risk. Though I'm not sure I agree […]. The EA Facebook group has 9,158 members; should we treat that as part of our "public face" and only discuss schistosomiasis and GiveDirectly there? I also buy that filtering for "people who can seriously examine ideas even when they're weird" might be more helpful early in EA's development than filtering for "people who can appreciate the case for AMF." If we were optimizing for having the right "public face" I think we'd be talking more about things that are in between malaria nets and AI on the obviousness/weirdness spectrum, like biosecurity and macroeconomic policy reform. (Bensinger, 2016)

The broad outreach strategy pursued by leading EA organizations in the period 2013-2016 led to people joining the movement believing it was about effective giving to evidence-backed charities. Some then became disillusioned or felt tricked as they engaged more deeply with the movement and were encouraged to shift to AI-safety/x-risk; it seemed that EA was engaging in a bait-and-switch tactic:

> I used to work for an organization in EA, and I am still quite active in the community. 1 - I've heard people say things like, "Sure, we say that effective altruism is about global poverty, but -- wink, nod -- that's just what we do to get people in the door so that we can convert them to helping out with AI/animal suffering/(insert weird cause here)." This disturbs me. (Anonymous#23, 2017; quote marks and parentheses in original)

> [I]n my time as a community builder […] I saw the downsides of this. I saw […] Concerns that the EA community is doing a bait-and-switch tactic of "come to us for resources on how to do good. Actually, the answer is this thing and we knew all along and were just pretending to be open to your thing". (User: weeatquince [Sam Hilton], 2020; quote marks in original)

> [M]y perception is that many x-risk people have been clear from the start that they view the rest of EA merely as a recruitment tool to get people interested in the concept and then convert them to Xrisk causes. (Pearce, 2015).

Given that this is how movement leaders have also portrayed the movement, it is not a stretch to find truth in these claims. Below is an excerpt from a 2011 discussion on *Felicifia*, between an influential GWWC leader/CEA affiliate (writing pseudonymously as 'Utilitymonster'[28]) and a high-school student (who would go on to hold influential positions in EA organizations and for EA Funders; their name is redacted here because they were a minor at the time) about 80K, then still called High Impact Careers:

> Utilitymonster: GWWC is explicitly poverty-focused but high impact careers (HIC) is not. In fact, hardcore members of GWWC are heavily interested in x-risk, and I estimate that 10-15% of its general membership is as well. I'd take them seriously as a group for promoting utilitarianism in general. I'm a GWWC leader.
>
> [Redacted]: but HIC always seems to talk about things in terms of "lives saved", ive never heard them mentioning other things to donate to. [...]
>
> Utilitymonster: That's exactly the right thing for HIC to do. Talk about lives saved with their public face, let hardcore members hear about x-risk, and then, in the future, if some excellent x-risk opportunity arises, direct resources to x-risk. (User: utilitymonster, 2011a, 2011b)

EA organizations now stress in various guides that while it is often better to start off using examples from global poverty and animal welfare when introducing someone to EA, one should avoid the perception of a bait-and-switch by mentioning that AI-safety/x-risk are also important cause areas for many EAs:

> Discussing lesser known cause areas such as AI safety or wild animal suffering can elicit skepticism [...]. It is likely easier, and possibly more compelling, to talk about cause areas that are more widely understood and cared about, such as global poverty and animal welfare. However, mentioning only one or two less controversial causes might be misleading, e.g. a person could become interested through evidence-based effective global poverty interventions, and feel misled at an EA event mostly discussing highly speculative research into a cause area they don't understand or care about. This can feel like a "bait and switch"—they are baited with something they care about and then the conversation is switched to another area. (EA Hub, 2019b)

With the rise of 'longtermism' and EA's openness about the movement's prioritization of AI-safety/x-risk, public-facing EA content now introduces the idea of existential or global catastrophes using careful, noncontroversial language (e.g., caring about future generations; believing all people matter) and highlighting examples that resonate with contemporary concerns and events, such as climate change, nuclear risks, and pandemics, then adding—almost as an afterthought—that the category of risks also includes emerging technologies like AI and biotechnology/synthetic biology.

Another issue that arose from EA's public-facing content was the *quality* of individuals being attracted to the movement, since casual participants failed to accurately grasp the nuances of the EA worldview:

> Oversimplified messaging can propagate misunderstandings of effective altruism, for example that it is just about saving lives, is focused on earning to give, or is only for utilitarians. [...] Over-simplified messaging may also attract people who don't share the community's goals, or who don't appreciate the complexity of the endeavour (Centre for Effective Altruism, 2020)
>
> An EA group that anyone can join and leave at a whim is going to have relatively low standards. This is fine for recruiting new people. But right now I think the most urgent EA needs have more do with getting people from the middle-of-the-funnel to the end, rather than the beginning-of-the-funnel to the middle. And I think helping the middle requires a higher expectation of effort and knowledge. (User: Raemon [Raymond Arnold], 2019)

[28] The criteria in my ethical checklist for attaching a 'real' name to a 'username' are not met for Utilitymonster, hence I cannot reveal their identity. What I *can* say is that it is clear from Utilitymonster's interactions with others on *Felicifia* that Utilitymonster is representing their affiliations honestly—in this post they state that they are a leader at GWWC and, in a later post, they also state they are affiliated with CEA (User: utilitymonster, 2012). Both Toby Ord and Will MacAskill (who were also leaders of GWWC and co-founders of CEA) posted on *Felicifia* in this period using their real names (MacAskill was then 'Crouch'), and neither they, nor other key movement figures who interacted on posts with Utilitymonster such as Pablo Stafforini, Ryan Carey, Brian Tomasik, Holly Morgan, Carl Shulman, and Ruiari Donnelly, ever questioned Utilitymonster's claims regarding their position and role in the community.

In other words, an EA group with a low threshold for entry may attract many people, but most of those people will not *really* be 'EA'. As summarized in the notes from a conversation between EA communications specialist Michael Bitton and prominent movement leader Peter Wildeford (then Hurford), *really* being 'EA' typically means being open to the fringier topics more closely associated with transhumanism:

> [T]hings in the ea community could be a turn-off to some people. While the connection to utilitarianism is ok, things like cryonics, transhumanism, insect suffering, AGI, eugenics, whole brain emulation, suffering subroutines, the cost-effectiveness of having kids, polyamory, intelligence-enhancing drugs, the ethics of terraforming, bioterrorism, nanotechnology, synthetic biology, mindhacking, etc. might not appeal well. There's a chance that people might accept the more mainstream global poverty angle, but be turned off by other aspects of EA. [...] Obviously, we want to attract more people, but also people that are more EA. (Bitton and Hurford in Hurford, 2014)

This led to emphasis on the **'fidelity model'** of spreading EA ideas, elaborated by Kerry Vaughan (then working at CEA on EA community-building). Vaughan argued that important EA ideas risked dilution and distortion when filtered through mass media and other mechanisms designed to reach a broad audience, much like the children's game of Telephone (wherein a message whispered along a chain of people becomes garbled, ending up hilariously different from what was originally said). EA, he explained,

> contains a large number of nuanced and interrelated ideas. Some methods of spreading these ideas require stripping away either the depth of the ideas or their nuance, or both. When the context gets stripped away, those who receive the ideas leave with something that's similar to effective altruism, but different. Thus, when we hear the EA message repeated back to us, we get sentences like "EA is about earning all the money you can and donating it to GiveWell charities" or "EAs only care about interventions that are supported by randomized controlled trials." (Vaughan, 2017)

It was more important, he wrote, to target fewer people and retain nuance, adding that '[p]eople seem to update better when talking to each other in person', thus one-on-one conversations should be prioritized (ibid.). Hence, rather than focusing on attracting a wide range of people at the top of the funnel, EA began to prioritize recruiting people who are already predisposed to the 'core' EA message and thus more likely to move all the way through the funnel to becoming a 'core' EA:

> [P]eople enter the funnel when they are first introduced to effective altruism, and exit either through leakage at any stage, or out of the bottom as deeply engaged individuals leading high-impact careers. Leakage should always be avoided, but prioritization *within* and *between* stages is very important [...] Trying to get a few people **all the way through** the funnel is more important than getting *every* person to the next stage. (EffectiveAltruism.org, 2019, emphases in original)

> Currently, we are focusing mostly on Step 4: Contributors to Core. We think that there are many contributors in the community who are willing and able to work full-time on some of the world's most pressing problems, and we are trying to help them build the skills and network to do so. (Centre for Effective Altruism, 2018b)

Recently, however, EAs have started to re-emphasize recruitment at the top of the funnel to ensure that the movement continues to grow:

> Going forward, we'll eventually need to get the top-of-funnel metrics growing again, or the stock of 'medium' engaged people will run out, and the number of 'highly' engaged people will stop growing. Outreach to new people seems to be getting more highly prioritised going forward. (Todd, 2021)

I address this further in the next section, wherein I describe how EA is beginning to re-prioritize broad outreach and effective giving as a strategy to maintain credibility.

5. Navigating EA's Identity as a Movement: A 'Tricky Balancing Act'

As shown in the preceding sections, EA leaders, organizations, and core community-members acknowledge that the movement maintains two versions of EA, although the trade-off between keeping EA broadly appealing versus attracting people who will be sympathetic to core EA can be challenging. MacAskill explains that defining EA is 'a matter of engineering,' requiring a 'tricky balancing act' to accurately describe 'the actual practice of those who are currently described as engaging in effective altruism, and the understanding of effective altruism that the leaders of the community have' on one hand, while also 'ensur[ing] that the concept has as much public value as possible' on the other (MacAskill, n.d., p. 2).

Initially, the benefits of public-facing EA were obvious: even critics of EA felt obliged to heap praise upon the movement's members. For example, when Giles Fraser debated MacAskill on *Intelligence Squared* in 2015, he prefaced his criticism with the following disclaimer:

> I don't like arguing with Will, because Will's a saint, and there's nothing to be had with, there's no upside, to arguing with someone as good and saintly and interesting as Will [...] you're just bloody brilliant, mate! (Fraser, in Intelligence Squared, 2015; at 00:17:12)

Maintaining a substantial number of grassroots effective altruists who continue to promote the idea that EA is just about effective giving and ending global poverty helps counter accusations that EA is mainly about paying EAs to work on AI-safety/x-risk. However, in light of the movement-wide embrace of 'longtermism' and the enormous sums of money flowing into the pockets of community-members, it has become increasingly difficult to reconcile public-facing EA with core EA's actions. The ongoing re-negotiation of EA identity is a form of 'ideological work', i.e., a remedial discourse employed by ideological groups to explain obvious disjunctions between their professed beliefs and their actual behavior (Snow, 2004, p. 398).

5.1. EA's Identity Crisis

What originally made 'Effective Altruism' seem unique and important was its focus on rigorous evidence and selfless behavior. Yet these days, most EAs endorse causes that are *not* backed by 'evidence and careful analysis' (as stipulated in the definition of EA) but by speculative reasoning and subjective expected value calculations, while EA's current spending habits can hardly be called altruistic — a reality widely acknowledged across the EA ecosystem, from peripheral community members/grassroots effective altruists who express *concern* about this spending (see, e.g., Beardsell, 2022; Rosenfeld, 2022) to the leadership figures who *endorse* this spending (see, e.g., MacAskill, 2022c; Vollmer, 2021). EAs are encouraged to work on AI safety, technical AI alignment research, AI/biotechnology policy, or biosecurity, but those who are not naturally talented in (or inclined toward) these areas are persuaded to work on EA movement-building and operations. Generous grants are awarded to EAs to develop new organizations popularizing 'longtermism'/x-risk, produce outreach content, including podcasts, YouTube videos, syllabi for university courses[29], courses and competitions for high school students[30],

[29] OpenPhil offers to pay university lecturers ~$30,000 - $50,000 to create and teach courses on EA/'longtermism' at their universities (Open Philanthropy, 2021d). For grants that have been awarded for university course development, see: (Open Philanthropy, 2018, 2021a, 2021b, 2021c, 2022d, 2022e, 2022f, 2022g, 2022h, 2022i, 2022j).

[30] To give an example of how swiftly teenagers are recruited and rewarded for their participation in EA: one 17-year old recounts how in the past year since they became involved in EA, they have gained some work experience at Bostrom's FHI; an internship at EA organization Charity Entrepreneurship; attended the EA summer program called the European Summer Program on Rationality (ESPR); been awarded an Open Philanthropy Undergraduate Scholarship (which offers full funding for an undergraduate degree); been awarded an Atlas Fellowship ($50,000 to be used for education, plus all-expenses paid summer program in the Bay Area); and received a grant of an undisclosed amount from CEA's EA Infrastructure Fund to drop out of high-school early, move to Oxford, and independently study for their A-Levels at the central EA hub, Trajan House, which houses CEA, FHI, and GPI among other EA organizations (*reference and individual's identifying features redacted because the individual is still under* 18; *link to reference can be provided upon request to readers seeking verification*). EA programs such as the **Atlas Fellowship** (Atlas Fellowship, 2022a, 2022b, 2022c, 2022d), **Leaf** (LEAF, 2021a, 2021b, 2021c), the **Summer Program on Applied Rationality** (Center for Applied Rationality, 2012; Davey, 2018), the **European Summer Program on Rationality** (ESPR, 2020), the **Apollo Fellowship** (Apollo Fellowship, 2022) are all-expenses paid programs designed start teenagers in the EA 'longtermism' career pipeline. Teenagers are also targeted by EA-funded contests that offer generous cash prizes, such as the **Eon Essay Contest** (Eon Essay Contest, 2022a, 2022b) which offers prizes of up to $15,000 to students who write a short essay about movement co-founder Toby Ord's latest book, *The Precipice* (thereby amplifying the attention given to Ord's book, particularly since participants can request a free copy), and the **Prometheus Science Bowl** competition

and even entire organizations and projects dedicated to pushing EA content in the media[31]. EAs receive funding to help them advance in EA-promoted careers (e.g., technical AI/biotechnology research/policymaking, EA movement-building and operations) and for personal wellbeing. This ranges from professional advancement (career coaching and job placement, thesis advisory service[32], funding advanced degrees, bespoke fellowships and internships at EA/EA-aligned institutes) to quite literally *anything* that improves the general wellbeing and productivity of individual EAs, such as funding living expenses, cleaning services, catering and meal provision, therapy and counseling[33], personal assistants, all-expenses-paid retreats[34] in idyllic locations (the movement has spent tens of millions of dollars purchasing opulent real estate, including two palatial estates that most people would feel comfortable calling 'castles', even if that is not the correct architectural term (User: SeeYouAnon, 2022), and at least four mansions)[35];

(Prometheus Science, 2022). OpenPhil recently provided funding to **Victory Briefs Institute** (which is not explicitly an EA organization but includes several EAs on staff) to create and include an EA curriculum in VBI's summer debate camp for high-schoolers (Open Philanthropy, 2022m).

[31] For example, Vox's **Future Perfect** is a donor-funded EA vertical that appears in a nationally renowned media outlet. Other examples include the **Longview Philanthropy (formerly: Effective Giving UK)/Effective Ideas *Blog Contest*** (Longview Philanthropy & Effective Ideas, 2022a, 2022b), in which 5 prizes of $100,000 each will be awarded next year for the 'best new and recent blogs' that promote 'longtermism' and EA *[note: this was funded by the FTX Foundation and the contest appears to be discontinued as of February* 2023]; the organization will also make grants to 'promising young writers' (Longview Philanthropy & Effective Ideas, 2022a). The **Existential Risk Observatory** is a Dutch EA project whose mission is to publish op-eds in the mainstream media about x-risk (Existential Risk Observatory, 2021, 2022). Another EA project, **Training for Good**, offers an EA Funds-funded **'Tarbell Fellowship'** which provides mentoring and pays $50,000 to early career journalists to write about EA themes (Training for Good, 2022).

[32] See (Effective Thesis, 2021).

[33] EA Funds provides grants to EA counselors/therapists to offer free/discounted sessions to EAs, and grants for individual EAs to seeking productivity coaching, therapy and health care (Centre for Effective Altruism, 2019b, 2019c, 2019d, 2021a). CEA even created a guide for non-EA therapists treating EA patients to understand community norms (Wise & Sasi, 2022). EA Funds encourages EAs to apply for grants for *personal* support, including to pay for health care (Bergal, 2021). This is not to suggest that health services are not important; however, as one commenter on the EA *Forum* points out, it not clear why funds donated by the general public (under the impression that the funds would support *the most effective* organizations working to improve the long-term future) are distributed as grants to fund health care of individual EA movement-members (Fritz, 2021). Another EA organization, Nonlinear, launched a Productivity Fund to increase productivity of EAs working in the 'longtermist' space by providing everything from mental health services to professional services like research assistants, personal assistants, tutors and products like SAD lamps, productivity apps/subscriptions, office supplies, electronics, etc. (Nonlinear, 2022). Another EA organization, **Rethink Wellbeing**, was recently created provide therapy and coaching for EAs: 'People who are well make others feel well. People who are well do good better. We aim to impact the present and future of humanity. That is why we tackle the Effective Altruism Community first. […] Our content is bespoke to the needs, problems, and preferences of effective altruists' (Rethink Wellbeing, 2023).

[34] The movement sponsors so many retreats that a new organization, **Canopy Retreats**, was created to provide logistical support to EA retreat organizers (Canopy Retreats, 2022).

[35] EA's current activity is centered on hosting retreats, workshops, and conferences for movement members so much so that EA leaders have justified spending over $44,000,000 buying properties to host such events. In 2017, the Center for Applied Rationality (CFAR, the sister organization of MIRI) purchased the $1.7 million dollar **Bay Hill Mansion** in Bodega Bay just north of San Francisco (Michaud, 2017) to host all-expenses paid workshops and retreats, such as the AI-Risk for Computer Scientists (AIRCS) workshop (Scholl, 2019; Telleen-Lawton, 2019; User: ArthurRainbow, 2020). In April 2022, CEA (now Effective Ventures) purchased **Wytham Abbey** (Axford, 2022; Lewis-Kraus, 2022) a manor house located on 25 acres of gardens and parkland just outside of Oxford, which was listed at £15,000,000 (Savills, 2021). A recent post on the EA Forum clarified that the purchase price 'was a bit under £15M' (Cotton-Barratt, 2022) and the recently-released title register for the property (HM Land Registry -GOV.UK, 2023) confirms the purchase price was £14,854,000 (approximately $19 million dollars). As of July 2022, the boutique architectural firm Jonathan Tuckey Design boasts that they 'are delighted to have been appointed to reimagine Wytham Abbey as a retreat for writers, charitable organisations and academic groups' (Jonathan Tuckey Design, 2022), suggesting that there is also a considerable budget for renovations; information about the grant for this purchase was released by OpenPhil in February 2023 shows that OpenPhil gave CEA (now Effective Ventures) $22,805,823 to fund the purchase and renovations (Open Philanthropy, 2021* [2023]). CEA also purchased the **Lakeside Guest House**, a B&B hotel set in a Victorian mansion at 118 Abingdon Road in Oxford, for £1,450,000 in June 2022 (HM Land Registry -GOV.UK, 2022). In November 2022, a Bay Area-based EA organization, Lightcone Infrastructure (the new name for team that maintains *LessWrong*) completed the purchase of the **Rose Garden Inn —**a hotel complex comprising two historical mansions and three contemporary buildings at 2740 and 2744 Telegraph Ave. and 2348 Ward St. in downtown Berkeley, CA (Habryka, 2022a; User: Raemon [Raymond Arnold], 2022), set on three connected land parcels (Alameda County Assessor's Office, 2022a, 2022b, 2022c)— for ~$16,500,000 with plans for ~$3,500,000 in renovations (Habryka, 2022b). The organizers of the European Summer Program in Rationality (ESPR; the European branch of CFAR's summer rationality program now spun off as a separate organization) and Czech EA, Jan Kulveit and Irena Kotikova, also recently completed the purchase of the **Chateau Hostacov** (Kotikova, 2022; Kulveit, 2022; Kurzy.cz, 2022; Štorkán, 2022), a castle located outside Prague, for ~$3,500,000 with an additional budget of ~$1,000,000 for other operational costs (Jacobs, 2022; Štorkán, 2022). These purchases are in addition to millions spent renting offices and co-working spaces for EAs and EA organizations: for instance, in

provision of housing and co-working spaces in desirable cities so that EAs can spend 'seasons' abroad in cities like Prague (Kotíková, 2022; Prague Fall Season, 2022), Mexico City (Malagon, 2022), or Cape Town (Pieters et al., 2022); and even a lavishly-funded 'fellowship' for EAs to live all-expenses-paid in the Bahamas for several months while working remotely on EA projects (FTX Foundation, 2021).

Other community-members have pointed out the 'tension within EA [...] between those who celebrate the increasing focus of highly-engaged EAs and major orgs on long-termism and those who embrace the older, short-termism focus on high standards of evidence' (Leong, 2022b). The identity crisis can be boiled down to a few circular questions, none presently resolved:

- Is EA currently (or, was it ever) 'effective altruism'?
- Can 'longtermist EA' and 'neartermist/effective giving EA' co-exist under the EA brand?
    - If so, how can they be reconciled as part of the same movement?
    - If not, which one continues using the 'EA' brand and which one becomes separate?
- Is it better to promote 'longtermism' or 'x-risk reduction' to make EA goals more appealing to the general public, academics, and policymakers?

As EA leaders and movement members grapple with these questions, solutions remain contested and are not permanent fixes so much as temporary stop-gaps. For example, EA thought-leaders often argue that critics of EA are being unreasonable and have misunderstood EA, since EA just means wanting to help others as much as possible and using evidence and analysis to take action to do so — an aim that is difficult to disagree with (Alexander, 2022a; Wiblin, 2015a). Yet contrast this against how EA thought-leaders and core community-members describe EA when engaging in (community-facing) debates about broad vs. targeted outreach; in these non-public-facing materials, being an EA—*really* being an EA—means being 'value-aligned' with a rather narrow set of ideals, beliefs, and actions, deviation from which is discouraged (Cremer, 2020): Kelsey Piper worries that important EA ideas 'might get watered down to be more palatable as they spread more widely' (Piper, 2018), CEA cautions that 'over-simplified messaging may also attract people who don't share the community's goals, or who don't appreciate the complexity of the endeavour' (Centre for Effective Altruism, 2020), Hurford (now Wildeford) argues that 'obviously, we want to attract more people, but also people that are more EA' (Hurford, 2014). 'We face a dilemma,' writes a long-time community-member, 'if everything counts as EA, then EA will lose its distinctiveness - but at the same time we don't want to come off as narrow minded' (Leong, 2022a). EA ends up employing a classic 'motte-and-bailey' (Hazell, 2022a): as soon as the genuinely-held but controversial position (longtermism/x-risk/AI-safety) comes under criticism, EAs retreat to portraying EA as merely the unobjectionable goal of 'doing good' and 'helping others'.

Still, community-members are fully aware that EA is not actually an open-ended question but a set of conclusions and specific cause areas:

> Everyone reasonably familiar with EA knows that AI safety, pandemic preparedness, animal welfare and global poverty are considered EA cause areas, whereas feminism, LGBT rights, wildlife conservation and dental hygiene aren't (User: nadavb, 2021)

2022 OpenPhil granted $5,318,000 to CEA for three years of support for CEA's forthcoming **Boston Biosecurity Hub**, which will also house MIT genetic engineering researcher Kevin Esvelt's Sculpting Evolution lab (Open Philanthropy, 2022b); $8,875,000 to CEA to provide five years of support to CEA's just-launched **Harvard Square EA co-working space** (Open Philanthropy, 2022c), which is distinct from the Biosecurity Hub and will occupy the entire fourth floor of the iconic Abbot Building (Wolf, 2022); $250,000 to the *student* EA/x-risk group at Harvard University (**the Harvard AI-Safety Team, HAIST**) to rent and refurbish an office space for one year (Open Philanthropy, 2022a); $445,000 to the **London EA Hub** to support the EA groups at LSE, King's College, Imperial College, and UCL and to run a co-working space in London for people working on EA projects (Open Philanthropy, 2022l).

> [I]f someone was very pro-Fair Trade and started saying EA was all about Fair Trade, this would not really be representative of what most EAs think (even if this person was convinced that Fair Trade was the best cause within the EA framework). Naturally, in a movement as large as the EA movement, there remains a diversity of viewpoints, but, nonetheless, I think it's fairly easy for experienced EAs to have a sense of what is a common EA view, and what is not (Savoie, 2018)

> [I]f you look at what the central EA individuals, organisations, and materials are promoting, you very quickly get the impression that, to misquote Henry Ford, "you can have any view you want, so long as it's longtermism". (Plant, 2022)

The continued presence of people in the movement who consider themselves 'EAs' but prioritize effective giving to global poverty and animal welfare charities is welcome only insofar as this helps preserve EA's public image as cause-neutral, worldview-diverse, and deeply committed to helping others. In their more candid moments, however, EAs are far less sanguine about people who prioritize alleviating global poverty. For example, Gregory Lewis —a long-time community-member and thought-leader, who is often mentioned in popular media accounts of the movement (e.g., Matthews, 2013; Ough, 2017; The Economist, 2018)— has expressed that global poverty-focused EAs are less clever, less informed, and less able to reason than EAs who focus on AI and the far future:

> It seems to me that more people move *from* global poverty to far future causes than people move in the opposite direction (I suspect, but am less sure, the same applies between animal welfare and the far future). It also seems to me that (with many exceptions) far future EAs are generally better informed and cleverer than global poverty EAs. (Lewis, 2016)

> [T]he most involved [people] in the EA community strongly skew towards the far future cause area in general and AI in particular. I think they do so, bluntly, because these people have better access to the balance of reason, which in fact favours these being the most important things to work on. (Lewis, 2018)

One EA *Forum* post, titled "Effective Altruism is No Longer the Right Name for the Movement" (User: ParthThaya, 2022) points toward another schism in the community, regarding whether the branding should focus on 'longtermism' or 'x-risk reduction' (Alexander, 2022b; Lifland, 2022; Nanda, 2022; User: AISafetyIsNotLongtermist, 2022). Influential community-members like Scott Alexander have weighed in, arguing that the 'longtermism' branding might actually be problematic, since most EAs believe AI x-risks are *imminent* – not something that will only affect the far future:

> if you're under ~50, unaligned AI might kill you and everyone you know. Not your great-great-(...)-great-grandchildren in the year 30,000 AD. Not even your children. You and everyone you know. (Alexander, 2022b)

Alexander concludes that unlike 'longtermism', 'the existential risk framework immediately identifies a compelling problem (you and everyone you know might die) without asking your listener to accept controversial philosophical assumptions' (Alexander, 2022b). Another influential community-member wrote a provocatively titled post, "Simplify EA pitches to: Holy Shit, X-risk!", which similarly argues that

> If you believe the key claims of "there is a >=1% chance of AI causing x-risk and >=0.1% chance of bio causing x-risk in my lifetime" this is enough to justify the core action relevant points of EA. This clearly matters under most reasonable moral views and the common discussion of longtermism, future generations and other details of moral philosophy in intro materials is an unnecessary distraction. (Nanda, 2022)

The debate about whether EA should use 'longtermism' or 'x-risk' inadvertently highlights how adopting complex (and controversial) philosophical stances operates as a red herring, distracting movement members and observers into analyzing the logical consistency of each position, rather than the concrete effects: though x-risk and 'longtermism' are presented as being different, they promote the same exact activities. Alexander admits as much, noting that although the two positions may, *in theory*, lead to slightly different conclusions, this is only 'in ways that rarely affect real practice' (Alexander, 2022b). This is not a surprising finding; it is the whole point. The resolution to the 'longtermism' vs. 'x-risk' debate does not really matter, since the point was never to create a coherent philosophical position and then follow wherever it leads, but to identify *whatever messaging works best in a given context* to produce the outcomes that movement founders, thought leaders, and funders wish to see. 'X-risk' and

'longtermism' are, by design, two sides of the same coin. MacAskill makes this clear in his reply on Alexander's (2022b) post:

> The main view I'm putting forward in this comment is "we should promote a diversity of memes that we believe, see which ones catch on, and mould the ones that are catching on so that they are vibrant and compelling (in ways we endorse)." These memes include both "existential risk" and "longtermism". [...] What meme is good for which people is highly dependent on the person and the context (e.g., the best framing to use in a back-and-forth conversation may be different from one in a viral tweet). This favours diversity; having a toolkit of memes that we can use depending on what's best in context. (MacAskill, 2022b; note that MacAskill is putting his own beliefs in quotation marks for effect, not actually quoting something someone else said)

5.2. Effective Giving and "Hard-to-Fake" Signals

The most prominent 'ideological work' taking place in EA at the present is re-emphasizing the value of effective giving and encouraging movement members to follow-through on the GWWC pledge (Freeman, 2022; Freeman, Parikh, et al., 2022; Freeman, Townsend, et al., 2022). Many EAs believe that 'direct work' in an EA cause area constitutes their contribution: the resource that they can dedicate to producing 'impact' is their time and intellectual energy. The result is that many EAs no longer donate significantly, or at all, to effective charities — a fact widely acknowledged by community leaders (Hazell, 2022b). Whereas EAs initially depicted themselves as living on graduate student budgets and rarely treating themselves to small luxuries like movies or dinner out so as to have more money for bednets and deworming pills , today EA funding not only pays for graduate studies but also for research assistants and personal assistants, and for basically anything that EAs can claim improves their productivity and career prospects (skeptical readers are encouraged to peruse the grants database at Effective Altruism Funds, 2023). As Ben Todd (the co-founder of 80K) writes, 'in the past we used frugality as a hard-to-fake signal of altruism, but that signal no longer works' (Todd, 2022).

The need to establish new 'hard-to-fake signals' of EAs' integrity and commitment to helping others has prompted a great deal of collective soul-searching in the community (Freeman, Townsend, et al., 2022). The general consensus seems to be that re-emphasizing a norm of donating to global poverty and animal welfare charities provides reputational benefits, which compensate for 'wasting' donations on these less-effective causes (as most 'neartermist' causes are now understood to be). Providing token support to GiveWell-recommended charities working on global health and development is important for maintaining good 'epistemics' and 'community health'. In other words, it is important for maintaining EA's reputation and for presenting an easily digestible entry point into EA. As articulated by one longtime movement member:

> I also worr[y] about what the impact will be if too many people stop focusing on poverty despite agreeing that existential risk is much more important. Firstly, I think that successes in global poverty will help establish our credibility. Everyone cares about poverty and if we are having successes in this area, people will respect us more, even if they are skeptical of our other projects. (Leong, 2017)

It is well-recognized in the community that global poverty alleviation provides an 'unobjectionable gateway drug' (User: Michael_PJ, 2022), i.e., a relatively non-controversial on-ramp into EA. Some EAs now argue that it is very important to make donations to the 'classic' EA charities, as 'symbolic donations' to offset accusations of hypocrisy (Daniel, 2019). During one discussion on the EA *Forum*, a community member suggested making large donations to GiveWell-recommended charities to head off criticism from outsiders, since 'it looks a bit suspicious if we conclude that the best way to have an impact is mostly to pay already privileged people high salaries' (Colbourn, 2022).

Oliver Habryka—who is influential in EA through his role as a fund manager for the LTFF, a grantmaker for the Survival and Flourishing Fund, and leading the *LessWrong*/Lightcone Infrastructure team—has stated that the only reason he believes EA should continue supporting non-longtermist efforts is to preserve the public's perception of the movement:

> To be clear, my primary reason for why EA shouldn't entirely focus on longtermism is because that would to some degree violate some implicit promises that the EA community has made to the external world. If that wasn't the case, I think it would indeed make sense to deprioritize basically all the non-longtermist things. (Habryka, 2021)

Put bluntly, many EAs appear to believe effective giving to GiveWell charities matters primarily because it strengthens EA's credibility: the benefits to the global poor are just silver lining. Even those who believe that 'longtermism' is *objectively* a more effective cause area still understand that the global poverty angle means people will be more accepting of 'longtermism':

> It makes me wonder if some of the "original EA norms", like donating a substantial proportion of income or becoming vegan, might still be quite important to build trust, even as they seem less important in the grand scheme of things (mostly, the increase in the proportion of people believing in longtermism). (Woodside, 2022b; parentheses in original)[36]

> Anecdotally, I know many EAs who both endorse long-termism and donate to AMF. In fact, my guess is that a majority of long-termist EAs donate to organizations that have been selected for their short-term benefits. As I say in another comment, I'm not sure this is a mistake because 'symbolic' considerations may outweigh attempts to directly maximize the impact of one's donations. (Daniel, 2019)[37]

While others, perhaps cognizant than openly endorsing the use of global health charities as a PR strategy is, itself, very poor PR, argue that continuing to support 'neartermist' or 'shortermist' causes helps maintain *good epistemics* in the community (Leong, 2022b).

5.3. Addressing 'Misconceptions' About EA

Another form of remedial discourse attempts to correct for the bait-and-switch dynamic by publishing **alternative histories of the movement**. This includes a historical revisionism, whereby leading EAs now say that evidence was never *actually* important in EA. During the January 2013 GiveWell Board meeting, in which the organization discussed their reasons for spinning off GiveWell Labs into Open Philanthropy, Karnofsky asserted that proven, evidence-based interventions had never *really* been part of GiveWell's vision:

> I do not agree with the attitude that is being expressed here, that, like, the way to do the most good is through LLIN[38] evidence. I've really never believed that. There were reasons that we focused on proven interventions for GiveWell initially, and I can go over what those were, but you know, I've really never believed that the way to accomplish the most good, that the only path to that, is through evidence. (Karnofsky in GiveWell, 2013a; starts at 00:22:37)

Ben Todd has similarly explained that his definition of EA had never *actually* been based on 'using evidence and reason':

> My definition didn't have "Using evidence and reason" actually as part of the fundamental definition. I'm just saying we should seek the best ways of helping others through whatever means are best to find those things. And obviously, I'm pretty keen on using evidence and reason, but I wouldn't foreground it. (Todd, in Koehler et al., 2020; at 00:21:02)

Such claims seem strange, since (1) GiveWell and GWWC built their reputations by arguing that people should only donate to charities proven to be the most cost-effective through evidence, (2) EA organizational websites and definitions prominently feature the phrase 'evidence and careful reasoning', and (3) EAs routinely stress the importance of evidence in public facing content— e.g., 'Effective altruism takes a different approach. We simply want to do the most good for the most people, and we use the evidence to help us do that. We don't want to trick or manipulate people. We simply say, "Here are the facts…"' (MacAskill, in Illing & MacAskill, 2018; quote marks in original).

The co-existence of these two diametrically-opposed stances toward 'evidence' amongst EA leaders exemplifies the notion of a *discursive reservoir of plausible deniability* that I intro-

[36] Woodside led the Yale EA chapter as an undergraduate and has held positions at two EA organizations focused on AI-safety: the Center for Human-Compatible AI (CHAI) at UC Berkeley and the Center for AI Safety (Woodside, 2022a), and helped launch the Machine Learning Safety Scholars Program, an AI-safety fellowship funded by the FTX Foundation (Woodside et al., 2022).
[37] Daniel is currently a senior program associate at OpenPhil. Previously, he was chief of staff at the Forethought Foundation (another EA organization started by Will MacAskill, based at Oxford), a research scholar at FHI, and the executive director of the Foundational Research Institute (FRI), a project of the EA Foundation.
[38] LLIN stands for 'long lasting insecticide-treated nets'; Karnofsky is referring to the evidence that GiveWell (and EA in general) uses to promote anti-malarial bed net distribution via the Against Malaria Foundation (AMF) as the most cost-effective opportunity to save lives [see, e.g. "Mass Distribution of Long-Lasting Insecticide-Treated Nets (LLINs)" (GiveWell, 2013b)].

duced in Section 3: when EAs needed to convince the public that they were credible experts pursuing a rigorous, scientific approach toward charitable giving, public-facing EA emphasized EA's commitment to evidence from RCTs. At the same time, EAs' inward-facing communications downplayed and even dismissed the value of evidence. Then, when scholars criticize EA for being too reliant on RCTs and quantitative evidence, EAs can defuse such criticism by referencing their (formerly) inward-facing communications as evidence that they do not believe RCTs or quantitative evidence should be a decisive factor in allocating one's resources.

6. Conclusions

EA has long been understood by the mainstream public, academics, and journalists to refer almost exclusively to the idea of effective giving, based on Peter Singer's efforts to alleviate global poverty. However, starting from its inception, the EA movement quietly and consistently advocated for what is now being promoted as 'longtermism' — a rebranding of the AI-safety/x-risk agenda developed by transhumanists in the early 2000s.

This article focused on how EA maintained these two different versions of EA at the same time, and can be distilled into the following points:

- EA discourse makes use of glittering generalities (e.g., 'doing good', 'cause neutrality', 'impartiality', 'openness to unusual ideas') to communicate on two different registers simultaneously — a double rhetoric consisting of public-facing and core content.

- Communication strategies such as the 'funnel model' have facilitated the enduring perception amongst the broader public, academics and journalists that 'EA' is synonymous with 'public-facing EA'. As a result, many people are confused by EA's seemingly sudden shift toward 'longtermism', particularly AI/x-risk; however, this 'shift' merely represents a shift in EA's *communication strategy* to more openly present the movement's core aims.

- The movement's founders and core leaders prioritized the x-risk/AI-safety agenda from the movement's inception, and considered global poverty alleviation merely a first step in converting new recruits toward prioritizing AI-safety/x-risk.

- The disjuncture between public-facing EA and core EA is rendered in sharp relief by the movement's current spending habits: money is lavished on EA community-members, justified by the 'longtermist' doctrine. Such spending contradicts the rhetoric of public-facing discourse, and has created an identity crisis for the EA movement: the movement clearly prioritizes 'longtermism'/AI-safety/x-risk, but still wishes to benefit from the credibility that global poverty-focused EA brings.

The purpose of clearly articulating and evidencing these points is to support the paper's larger aim: correcting the enduring (mis)perception amongst academics, journalists, and the general public that EA's main focus is global poverty reduction. Given the enormous amount of public-facing EA content, academics and journalists who start from the assumption that EA is about global poverty can peruse EA content and find abundant material that confirms that prior belief. This article shows that a more inductive approach to EA, which does not start from any assumptions but seeks to apprehend the movement as it actually exists, renders EA legible as a much different project.

As already mentioned, the present article represents just one chapter extracted from a book-length forthcoming study about EA. I had two choices when presented with the opportunity to write an article based on this forthcoming study: a broad but shallow overview summarizing the most salient points, or a narrow but deep analysis of one facet of EA. I opted for the latter, aiming to provide a fine-grained, ethnographically-informed academic study of EA

qua the EA movement by focusing on EA's public-facing versus core communication strategy. There is much that I do not address in this paper; for instance, I did not delve into what the AI-safety/x-risk agenda actually consists of, nor the actual concrete activities that EA's 'longtermist' agenda set in motion, in terms of promoting a very specific vision of the technological future while serving private sector interests. I do not deny the importance of investigating those links; rather, the present article is intended as a starting point for engaging with those larger issues.

Too often, academics paint EA with broad brushstrokes based on public-facing content, missing what EA—*actually existing* EA, the EA practiced by the movement's founders, leaders, funders, and core members—is really doing. The 'EA' that academics write about is a mirage, albeit one invoked as shorthand for a very real phenomenon, i.e., the elevation of RCTs and quantitative evaluation methods in the aid and development sector. This does not preclude the articulation of extremely valuable articles being written about that important topic. Rather, my point is that these articles and the arguments they make—sophisticated and valuable as they are—are not about EA: they are about the Singer-solution to global poverty, effective giving, and about the role of RCTs and quantitative evaluation methods in development practice. EA is an entirely different project, and the magnitude and implications of that project cannot be grasped until people are willing to look at the evidence beyond EA's glossy front-cover, and see what activities and aims the EA movement *actually* prioritizes, how funding is *actually* distributed, whose agenda is *actually* pursued, and whose interests are *actually* served.

Appendix A: Forums/Blogs Consulted as Primary Sources

Forums/Group blogs	Accelerating Future AI Impacts Alignment Forum EA Forum Extropians Felicifia LessWrong Overcoming Bias Oxford Prioritisation Project SL4	
Individual blogs	(Old) Felicifia [Seth Baum]	KelseyLikes [Kelsey Piper]
	A Mirror Clear [Toby Ord]	LuciusCaviola.com
	A Nice Place to Live [Michael Bitton]	LukeMuehlhauser.com
	A Path That's Clear	MaartenBoudry.blogspot.com
	Abolitionist [David Pearce]	Mcntyr.com [Peter McIntyre]
	Acritch.com [Andrew Critch]	Measuring Shadows [Sam Bankman-Fried]
	Aisafety.wordpress.com [Michael Cohen]	Meteuphoric [Katja Grace]
	Alexander Kruel	Michaeldello.com [Michael Dello-Iacovo]
	Aron's Blog [Aron Vallinder]	MilesBrundage.com
	Astral Codex Ten [Scott Alexander]	Minding Our Way [Nate Soares]
	Back of the Envelope Guide [Gordon Irlam]	Multiheaded1793tumblr
	Bayesian Investor [Peter McCluskey]	NickBeckstead.com
	BenKuhn.net	NickBostrom.com
	BenLandauTaylor.com	Nintil [Jose Luis Ricon]
	Black Belt Bayesian [Steve Rayhawk]	OldBlog.RobWiblin [Rob Wiblin]
	BrianTomasik.com	Otium [Sarah Constantin]
	careyryan.com [Ryan Carey]	Pablo's Miscellany [Pablo Stafforini]
	Common Sense Atheism [Luke Muehlhauser]	PaulChristiano.com
	Compass Rose [Ben Hoffman]	Philosophical Multicore [Michael Dickens]
	CullenOKeefe.com	Philosophy, et cetera [Richard Yetter Chappell]
	DanielDewey.net	Put a Number on It [Jacob Falkovich]
	DavidManley.squarespace.com	Rational Altruist [Paul Christiano]
	DeigoCaleiro.com	Rational Ethicist [Stijn Bruers]
	Don't Worry About the Vase [Zvi Mowshowitz]	Rational Futurist [Thomas McCabe]
	Duncan Sabien - Medium	Rationalist Conspiracy [Alyssa Vance]
	EA Coaching [Lynette Bye]	Reflective Disequilibrium [Carl Shulman]
	EricGastfriend.com	RobWiblin.com
	Everyday Utilitarian [Peter Hurford/Wildeford]	Samo Burja - Medium
	Flight From Perfection [Milan Griffes]	SamoBurja.com
	Fragile Credences [Tom Sittler-Adamczewski]	Sarah Constantin PostHaven
	Giving Gladly [Julia Wise]	SebastianFarquhar.com
	GreatPlay.net [Peter Hurford/Wildeford]	SethBaum.com
	Guzey.com [Alexey Guzey]	Shlegeris.com [Buck Shlegeris]
	Holly Elmore	SimonKnutsson.com
	Immortality Roadmap [Alex Turchin]	sites.google.com/site/nbeckstead [Nick Beckstead]
	Impartial Priorities [Denis Drescher]	Slate Star Codex [Scott Alexander]
	JeffreyLadish.com	Strucsurety [Connor Flexman]
	jefftk.com [Jeff Kaufman]	TeganMccaslin.wordpress.com
	JGMatheny.org [Jason Gaverick Matheny]	The Ne Plus Ultra Project [Pablo Stafforini]
	Guzey.com [Alexey Guzey]	The Unit of Caring [Kelsey Piper]
	Immortality Roadmap [Alex Turchin]	Thing of Things [Ozy Brennan]
	Impartial Priorities [Denis Drescher]	TobyOrd.com
	JeffreyLadish.com	Topher Brennan - Medium [Christopher Brennan]
	jefftk.com [Jeff Kaufman]	Topher Hallquist - Wordpress [Christopher Hallquist]

	JGMatheny.org [Jason Gaverick Matheny]	Uncredible Hallq [Christopher Hallquist]
	JohnSalvatier.org	Unstable Ontology [Jessica Taylor]
	JonahSinick.com	Update Project [Julia Galef]
	JonasMueller.net	Utilitarian Essays [Brian Tomasik]
	jsalvatier.wordpress	xuenay.livejournal.com [Kaj Sotala]
	KatjaGrace.com	Yudkowsky.net [Eliezer Yudkowsky]

Appendix B: EA Organizations

Overview of EA Organizations/Entities Operating in EA Ecosystem

Name	Purpose	Year founded
1 Day Sooner	Pandemic preparedness/biotech	2020
80,000 Hours	EA community support (careers)	2011
A Happier World	Media - videos (longtermism)	2020
ACE	Effective giving (animal welfare)	2013
ACX Grants	Longtermism/biotech regranter	2022
Adapt Research	Survival & resilience	2015
AI Governance Summer Program	AI governance	2022
AI Impacts	AI safety	2014
AI Objectives Research Institute	AI safety	2021
AI Safety Camp	AI safety	2018
AI Safety Field Building Hub (Vael Gates)	AI safety - community support	2022
AI Safety Hub (Julia Karbing)	AI safety – community support	2022
AI Safety Research Program	AI safety	2019
AI Safety Support	health support for AI safety researchers	2020
AI X-risk Podcast (AXRP)	Media (longtermism)	2020
AGI Safety Fundamentals	AI safety - course	2022
ALERT (Active Longtermist Emergency Response Team)	longtermism	2022
Aligned AI	AI safety	2022
Alignment Research Center (ARC)	Technical AI Research	2021
AllFed	Survival & resilience	2017
altLabs	biotechnology	2019
Altruistic Agency (Markus Magnuson)	EA operations - tech support	2022
Alvea / Telis Bioscience	Biotech, pandemic preparedness	2022
Anthropic	Technical AI research	2020
Anti Entropy	EA operations support	2022
Apart Research/AI Safety Ideas	Technical AI research	2022
APPG-Future Gens. / High Impact Policy Engine (HIPE)	Longtermism/policy	2017
Arb Research	Forecasting	2022
Assoc. for Long Term Existence & Resilience (ALTER)	Longtermism	2021
Asterisk (digital magazine)	EA media	2022
Atlas Fellowship	Longtermism - teenagers	2022
Better Matters (digital magazine)	EA media	2022
Berkeley Existential Risk Initiative (BERI)	x-risk/AI-safety support (funding conduit)	2017
Berkeley REACH	EA Community support	2016
Blue Dot Impact	AI Safety/x-risk/biotech career dev.	2022
Californians Against Pandemics	Pandemic Preparedness	2021
Cambridge AI Safety Hub	AI-safety	2022
Cambridge Boston Alignment Initiative	AI-safety	2022
Cambridge Existential Risk Initiative (CERI)	x-risk/longtermism, AI-safety	2022
CampusPA	Personal assistants	2020
Canopy Retreats	EA operations support (logistics)	2022
CaSPAR	Rationality training - teenagers	2020
Center for AI Safety	AI safety	2022
Center for Alcohol Policy Solutions	Global health and wellbeing	2021
Center for Applied Rationality	Rationality, AI safety	2012
Center for Election Science	Approval voting	2011
Center for Emerging Risk Research (CERR, now Polaris)	AI safety - Regranting & support	2019
Center for Human-Compatible AI (Berkeley CHAI)	Technical AI research	2015
Center for Reducing Suffering	Longtermism (suffering-focused)	2020
Center for Security and Emerging Technology (CSET)	AI/emerging technology policy/lobbying	2019

Center for Space Governance	Space governance	2022
Center on Long-Term Risk (formerly FRI)	Longtermism	2014
Centre for the Governance of AI (GovAI)	AI governance	2016
Centre for Effective Altruism (CEA)/Effective Ventures	EA Community building	2012
Centre for Long Term Resilience (form. Alpenglow)	Longtermism/policy	2019
České priority	Policy influence	2022
Charity Entrepreneurship	Charity incubation	2016
Charity Science	Effective Giving	2014
Charter Cities Institute	Charter cities, policy	2018
Clearer Thinking	Rationality/community-building	2018
Cognito Mentoring	Rationality	2013
Concordia Consulting (Brian Tse, Ashley Lin)	AI/emerging technology governance/policy	2021
Confido	Forecasting/rationality	2022
Conjecture	Technical AI research	2022
Constellation	EA community support (logistics)	2021
Convergence Analysis	Longtermism	2019
Convergent Research	Biotechnology	2022
Cooperative AI Foundation	Technical AI research	2020
Center for the Study of Existential Risk	x-risk/global catstrophic risks	2012
Center for Security and Emerging Technology	AI/biotech governance, policy	2019
Differential Projections	Forecasting	2022
Disputas	EA organization	2020
EA Coaching (now Lynette Bye Coaching)	EA community support (careers)	2017
EA Consulting	EA community support (careers)	?
EA Debate championship	Longtermism - teenagers	?
EA Foundation	Longtermism/AI-safety, suffering	2015
EA Funds	Funder	2017
EA Good Governance Project	EA community support (headhunting)	2023
EA Groups Resource Centre	EA community support (guides, templates)	2022
EA Hotel/CEALAR	EA community support (housing)	2017
EA Market Testing	EA promotion	2021
EA Medicine	EA promotion/community	2022
EA Mental Health Navigator *(incorporated 2022)*	EA community support	2019
EA Outreach	EA promotion	2015
EA Pathfinder	EA community support (careers)	2022
EA Radio	Media - podcast	2014
EA Summer Communications Fellowship	EA promotion	2022
East Bay Biosecurity	Biotechnology	2018
Effective Altruism Psychology Lab (NYU)	EA research	2022
Effective Environmentalism	Effective giving	2020
Effective Giving (NL)	Effective giving/longtermism	2018
Effective Giving (UK) now Longview Philanthropy	Longtermism, funder	2018
Effective Giving Quest	Effective giving	2021
Effective Self-Help	EA community support	2022
Effective Thesis	EA community support	2018
EffiSciences	Longtermism	2022
Emerging Effective Charity Evaluators	Effective giving	2020
Encultured AI	Technical AI alignment research	2022
Eon Essay Contest	Longtermism promotion - teenagers	2022
Epidemic Forecasting	Pandemic Preparedness	2021
Epoch	AI and forecasting	2022
European Summer Program in Rationality (ESPR)	AI Safety, rationality - teenagers	2017
Existential Risk Alliance (ERA – spin off of CERI)	X-risk, longtermism	2022
Existential Risk Observatory	Media - longtermism promotion	2021
Fonix Bioweapons Shelter Project	Survival & resilience	2022
Forecasting Research Institute	Prediction and forecasting	2022
Forethought	longtermism	2018

Foundations of Cooperative AI Lab	Technical AI research	2022
Founders Pledge	Effective Giving	2015
FTX Foundation	Funder	2021
FTX Future Fund	longtermism	2022
Fund for Alignment Research (FAR)	AI safety	2022
Future Forum	conference	2022
Future of Humanity Foundation	Funder, longtermism	2020
Future of Humanity Institute	X-risk, longtermism, AI safety	2005
Future of Life Institute	X-risk, longtermism, AI safety	2015
Future Matters Project	Policy influence	2022
GCR Policy	Longtermism policymaking	2022
Generation Pledge	Effective Giving	2019
GiveWell	Effective Giving	2007
Giving Alpha	Effective Giving	2018
Giving What We Can	Effective Giving	2009
Global Catastrophic Risk Institute	x-risk/global catastrophic risks	2011
Global Challenges Project	EA organization	2021
Global Priorities Encyclopedia	Prioritization/longtermism	2022
Global Priorities Institute	Longtermism, prioritization	2018
Good Ancestor Project	Longtermism promotin	2022
Good Forever	Regranter, longtermism/biotech	2022
Good Policies	Tobacco control	2019
Good Ventures (see Open Philanthropy)	Funder	2010
Guarding Against Pandemics	Pandemic Preparedness, political action	2021
Happier Lives Institute	global health and wellbeing	2018
Harvard AI Safety Team (HAIST)	AI Safety	2022
Harvard-MIT X-Risk	AI safety	2022
Hear This Idea	Media - longtermism podcast	2019
Herbivorize Predators	Wild animal suffering/biotechnology	2021
High Impact Athletes	Effective Giving	2020
High Impact Medicine	EA promotion (medical professionals)	2021
High Impact Policy Engine (HIPE)	EA promotion (policymakers)	2019
High Impact Professionals	EA community support (careers)	2021
Hofvarpnir AI	AI safety community support	2022
Human-Aligned AI Summer School	AI safety outreach	2018
Ideas Machines	Media - podcast	2022
Impact Markets	Impact market	2022
The Insect Institute	Insect welfare	2023
Institute for Progress	Progress studies, emerging tech policy	2022
Invincible Wellbeing	biotechnology	2021
Language Model Safety Fund	Technical AI research	2021
Lantern Ventures	Cryptocurrency	2018
Leaf	Longtermism course - teenagers	2021
LEEP	Global health and wellbeing	2021
Legal Priorities Project	Prioritization/longtermism (legal)	2019
Leverage	Rationality	2011
Lightcone Infrastructure	Community support (logistics)	2021
London Existential Risk Initiative	AI safety, x-risk	2022
Longtermism Fund	Longtermism, funder/regranter	2022
Longtermist Incubator (Jade Leung)	Longtermism organizational support	2020
Magnify Mentoring (form. WANBAM)	EA community support (careers)	2020
Manifold Markets	Prediction and forecasting	2022
Median Group	Longtermism/AI safety	2018
Metaculus	Prediction and forecasting	2015
Metaforecast	Prediction and forecasting	2022
MIRI	AI safety	2000
ML Safety	AI Safety	2022
Narratives	Media - podcast	2019

New York EA Hub	EA community support (co-working)	2022
NonLinear	AI safety, longtermism	2021
Non-Trivial Pursuits	EA career support - teenagers	2022
One for the World	Effective Giving	2014
Open Philanthropy	Funder	2014
OpenAI	Technical AI Research	2016
Organization for Preventing Intense Suffering	Biotech, cluster headaches/pain relief	2017
Ought	Technical AI research	2018
Pandemic Prevention Network	Pandemic preparedness, policy	2022
Parallel Forecasting	Forecasting	2018
People for the Ethical Treatment of Reinforcement Learners	Digital sentience welfare	2015
Phenomenological AI Safety Research Institute	AI safety	?
PIBBSS Summer Research Fellow	AI safety/longtermism	2022
Pineapple Operations	EA community support (personal assists.)	2022
Probably Good	EA community support (careers)	2020
Prometheus Science bowl	Longtermism/x-risk - teenagers	2022
Protect Our Future	Pandemic Preparedness, political action	2021
Qualia Research Institute (QRI)	Neurotechnology, consciousness	2018
Quantified Uncertainty Research Institute	Longtermism	2019
Radio Bostrom	Media - x-risk/podcast *about* Nick Bostrom	2022
Rational Animations	Longtermism	2021
Ready Research	Longtermism, effective giving	2019
Recursion Pharmaceuticals	Biotechnology	2013
Redwood Research	Technical AI Research	2021
Reserve stablecoin	Cryptocurrency	2019
Rethink Charity	Effective giving	2013
Rethink Priorities	Longtermism, survival/resilience, animals	2018
Rethink Wellbeing	EA community support (mental health)	2023
Riesgos Catastroficos Globales	x-risk/global catstrophic risks	2021
Road to AI Safety Excellence (RAISE)	AI Safety	2018
Roam Research	Note taking app	2019
Robert Miles AI Safety Channel	AI safety	2019
School of Thinking	Rationality	2021
Sentience Institute	Animal welfare, consciousness	2018
Sentience Politics	Animal welfare, consciousness	2013
Sentience Research	Neurotechnology, consciousness	?
Should We Studio	Media - videos	2021
Simon Institute for Long Term Governance	x-risk/global cat.risk., policy	2021
Social Change Lab	Social movement research	2022
Social Science Prediction	Prediction and forecasting	2019
SoGive	Effective giving, regranting	2017
Space Futures Initiative	Space governance	2022
Stanford Existential Risk Initiative (SERI)	AI safety, longtermism	2020
Summer Program on Applied Rationality and Cognition	AI safety, rationality - teenagers	2014
Survival and Flourishing Fund	AI safety, longtermism, progress studies	2019
Swift Centre	Prediction and forecasting	2022
Swiss Existential Risk Initiative (CHERI)	AI safety, longtermism, x-risk	2021
Tarbell Fellowship	Media - EA journalism prize	2022
tEAmwork (coworking space, Berlin)	EA community support (co-working)	2021
Telis Bioscience	Biotechnology	2019
The Life You Can Save	Effective giving	2012
ThinkBetter	Rationality	2019
Topos Institute	Math and technology research	2019
Utilitarian Podcast	Media - podcast	2021
Vox – Future Perfect	EA Media	2018
Wild Animal Initiative	Wild animal suffering	2019
Zeitgeist	Prediction markets/Futarchy	2021

References

Primary Sources

80000 Hours. (2012). *Causes - Future of Humanity Institute*. Retrieved May 1 from https://web.archive.org/web/20120501224911/http://80000hours.org/causes/future-of-humanity-institute

80000 Hours. (2016a). AI Safety syllabus. *80,000 Hours*. https://80000hours.org/articles/ai-safety-syllabus/

80000 Hours. (2016b). What are the world's biggest problems? [archived: 6/27/2016]. https://web.archive.org/web/20160627024429/https://80000hours.org/articles/cause-selection/

80000 Hours. (2017). No-one knew the world's worst problem, so we spent 8 years trying to find it [archived 3/30/2017]. *80000 Hours Career Guide*. https://web.archive.org/web/20170330065954/https://80000hours.org/career-guide/world-problems/

80000 Hours. (2021). *Our current list of pressing world problems [archived 10/27/2021]*. https://web.archive.org/web/20211027013526/https://80000hours.org/problem-profiles/

80000 Hours. (2022a). The highest-impact career paths our research has identified so far [archived 1/9/2022]. https://web.archive.org/web/20220109063513/https://80000hours.org/career-reviews/

80000 Hours. (2022b). *The highest impact career paths our research has identified so far [archived: 5/30/2022]*. https://web.archive.org/web/20220530165549/https://80000hours.org/career-reviews/

Agarwalla, V. (2019). *A Guide to Early-Stage EA Group-Building at Liberal Arts Colleges (Google Doc)*. https://docs.google.com/document/d/1F7Rjx6nxBnHEqqJxO-p0-8z0fPikFhiLhJRAspP0Z7w/

Alameda County Assessor's Office. (2022a). *Parcel Details for Parcel 54-1716-2*. https://propinfo.acgov.org/?PRINT_PARCEL=54-1716-2

Alameda County Assessor's Office. (2022b). *Parcel Details for Parcel 54-1716-3*. https://propinfo.acgov.org/?PRINT_PARCEL=54-1716-3

Alameda County Assessor's Office. (2022c). *Parcel Details for Parcel 54-1716-31*. https://propinfo.acgov.org/?PRINT_PARCEL=54-1716-31

Alexander, S. (2012, May 19). Dead children currency. *80,000 Hours*. https://web.archive.org/web/20190429025754/https://80000hours.org/2012/06/dead-children-currency-51/

Alexander, S. (2013 [2010]). Efficient Charity: Do Unto Others... *EA Forum*. https://web.archive.org/web/20190314152710/https://forum.effectivealtruism.org/posts/2iAwiBQm535ZSYSmo/efficient-charity-do-unto-others

Alexander, S. (2015a). Beware Systemic Change. *Slate Star Codex*. https://slatestarcodex.com/2015/09/22/beware-systemic-change/ , http://archive.is/w59X9

Alexander, S. (2015b, July 23). Stop Adding Zeros. *Slate Star Codex*. https://web.archive.org/web/20150813153432/http://slatestarcodex.com/2015/08/12/stop-adding-zeroes/

Alexander, S. (2022a). Effective Altruism as a Tower of Assumptions. *Astral Codex Ten*. https://web.archive.org/web/20220908090206/https://astralcodexten.substack.com/p/effective-altruism-as-a-tower-of

Alexander, S. (2022b). "Long-Termism" vs. "Existential Risk". *EA Forum*. https://web.archive.org/web/20220911050447/https://forum.effectivealtruism.org/posts/KDjEogAqWNTdddF9g/long-termism-vs-existential-risk

Alterman, T. (n.d.). *The EA Pitch Guide*. https://docs.google.com/document/d/1vsQdWIcL1nWdTTdQtB4uH1f_rIjDo27-CwaZUnfqEG4/edit

Anonymous#23. (2017). Anonymous EA comments (compiled by Rob Bensinger). *EA Forum*. https://forum.effectivealtruism.org/posts/QYGNmzqjhHgxfbyhy/anonymous-ea-comments?commentId=ddc7LhQyMKdXSqAio

Apollo Fellowship. (2022). *Apollo Fellowship - home page [archived 9/29/2022]*. https://web.archive.org/web/20220929090809/https://www.apollofellowship.com/

Atlas Fellowship. (2022a). *Altas Fellowship - Summer Program [archived 4/2/2022]*. https://archive.is/veUQ8

Atlas Fellowship. (2022b). *FAQ [archived 4/2/2022]*. https://web.archive.org/web/20220402145601/https://www.atlasfellowship.org/faq

Atlas Fellowship. (2022c). Fellowship [archived 4/2/2022]. https://archive.is/EVA1J

Atlas Fellowship. (2022d). *Welcome to the Atlas Fellowship [archived 4/2/2022]*. https://archive.is/pjrkq

Axford, J. (2022). Reply to question on: The Operations team at CEA transforms. *EA Forum*. https://forum.effectivealtruism.org/posts/hp2FWKhWiCto6oBrL/the-operations-team-at-cea-transforms?commentId=Dky2zoRTHDqaMcB4h#2Rj7mheX3bLHeMNd3

Bankman-Fried, S. (2012, July 19). Being a Utilitarian, Part 2: Conventional Charities. *Measuring Shadows*. https://web.archive.org/web/20170824020454/http://measuringshadowsblog.blogspot.com/2012/12/being-utilitarian-part-2-conventional.html

Beardsell, P. (2022). Comment on: Free-spending EA might be a big problem for optics and epistemics. *EA Forum*. https://forum.effectivealtruism.org/posts/HWaH8tNdsgEwNZu8B/free-spending-ea-might-be-a-big-problem-for-optics-and?commentId=FtdYrpmLfaEdF5ekx

Beckstead, N. (2010, June 29). Comment on: Singularity Summit. *GiveWell blog*. https://blog.givewell.org/2010/06/29/singularity-summit/#comment-141614

Beckstead, N. (2011a). Comment on: Maximizing cost-effectiveness via critical inquiry. https://blog.givewell.org/2011/11/10/maximizing-cost-effectiveness-via-critical-inquiry/#comment-244913

Beckstead, N. (2011b). Global Priority Setting and Existential Risk Crucial Ethical Considerations. *Future of Humanity Institute*. http://www.fhi.ox.ac.uk/wp-content/uploads/Global-priority-setting-and-existential-risk-crucial-ethical-considerations.pdf ; https://web.archive.org/web/20140409031741/http://www.fhi.ox.ac.uk/wp-content/uploads/Global-priority-setting-and-existential-risk-crucial-ethical-considerations.pdf

Beckstead, N. (2013). *On the Overwhelming Importance of Shaping the Far Future* [PhD, Philosophy, Rutgers University]. New Brunswick.

Beckstead, N., & Lee, M. (2011). Bold Giver Story: Nick Beckstead and Mark Lee. *Bolder Giving*. https://boldergiving.org/stories.php?story=Nick-Beckstead-and-Mark-Lee_105

Beckstead, N., Singer, P., & Wage, M. (2013). Preventing human extinction. *EA Forum*. https://forum.effectivealtruism.org/posts/tXoE6wrEQv7GoDivb/preventing-human-extinction

Bensinger, R. (2016). Comment on: Concerning MIRI's place in the EA movement. *Thing of Things*. https://archive.ph/94UTI#selection-1437.0-1441.435

Bergal, A. (2021). Reply to comment on: Public reports are now optional for EA Funds grantees. *EA Forum*. https://archive.is/rPfju

Bostrom, N. (1999). *What is transhumanism? Version 3.2 [archived 10/12/1999]*. https://web.archive.org/web/19991012083706/http://transhumanism.com/transhumanism.htm

Bostrom, N. (2002). Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards. *Journal of Evolution and Technology*, 9. https://www.nickbostrom.com/existential/risks.html

Bostrom, N. (2003). Astronomical Waste. https://www.nickbostrom.com/astronomical/waste.pdf

Bostrom, N. (2005a). A History of Transhumanist Thought. *Journal of Evolution and Technology*, 14(April).

Bostrom, N. (2005b). *Humanity's biggest problems aren't what you think they are (July 2005 Talk)*. https://www.youtube.com/watch?v=Yd9cf_vLviI; https://web.archive.org/web/20071126103931/http://www.ted.com/index.php/talks/view/id/44

Bostrom, N. (2013). Existential Risk Prevention as Global Priority. *Global Policy*, 4(1), 15–31.

Bostrom, N., & Ord, T. (2004). *Status Quo Bias in Bioethics: The Case for Human Enhancement* https://web.archive.org/web/20041220215352/http://www.nickbostrom.com/ethics/statusquo.pdf

Boudry, M. (2017, January 17). Think before you give: Charity is about a cool head as much as a warm heart. https://maartenboudry.blogspot.com/2017/07/think-before-you-give-charity-is-about.html

Brennan, O. (2017, May 18). Against EA PR. *Thing of Things*. https://thingofthings.wordpress.com/2017/09/20/against-ea-pr/

Canopy Retreats. (2022). *Canopy Retreats - About Us [archived 6/27/2022]*. https://web.archive.org/web/20220627175212/https://www.canopyretreats.org/about-us

Center for Applied Rationality. (2012). SPARC - Summer Program on Applied Rationality and Cognition. https://web.archive.org/web/20120422012442/http://appliedrationality.com/

Centre for Effective Altruism. (2017a, December 18). *CEA's 2017 Review and 2018 Plans*. Retrieved May 11 from https://web.archive.org/web/20190408003741/https://www.centreforeffectivealtruism.org/blog/cea-s-2017-review-and-2018-plans/

Centre for Effective Altruism. (2017b). *CEA's Guiding Principles [archive, 8/08/2018]*. https://web.archive.org/web/20180808144338/https://www.centreforeffectivealtruism.org/ceas-guiding-principles

Centre for Effective Altruism. (2017c). *CEA Fundraising Prospectus - Winter 2017*. https://assets.ctfassets.net/es8pp29e1wp8/heqArICf2SuEuMQmygEug/1051221345880fdcfc7562d3291b3b76/Centre_for_Effective_Altruism_____Fundraising_Prospectus_Winter_2017.pdf

Centre for Effective Altruism. (2017d). *Centre for Effective Altruism: Fundraising Prospectus - Winter 2017*. https://web.archive.org/web/20200220215856/https://assets.ctfassets.net/es8pp29e1wp8/heqArICf2SuEuMQmygEug/1051221345880fdcfc7562d3291b3b76/Centre_for_Effective_Altruism_____Fundraising_Prospectus_Winter_2017.pdf

Centre for Effective Altruism. (2018a, May). *CEA's Current Thinking [archived 11/29/2018]*. Retrieved May 12 from https://web.archive.org/web/20181129224534/https://www.centreforeffectivealtruism.org/ceas-current-thinking/

Centre for Effective Altruism. (2018b, May 12). The Funnel Model. https://web.archive.org/web/20190429035527/https://www.centreforeffectivealtruism.org/the-funnel-model

Centre for Effective Altruism. (2019a). *The Concentric Circles model*. https://www.centreforeffectivealtruism.org/the-concentric-circles-model

Centre for Effective Altruism. (2019b). *Fund Payout Report - November 2019: Long-Term Future Fund Grants*. https://archive.ph/wv3xj

Centre for Effective Altruism. (2019c). Payout Report - August 2019 Long-Term Future Fund Grants and Recommendations (updated version). *Effective Altruism Funds*. https://archive.fo/9i194

Centre for Effective Altruism. (2019d, February 27). Payout Report: August 2019 - Long Term Future Fund. *Effective Altruism Funds*. http://archive.is/9i194

Centre for Effective Altruism. (2020). *Effective Altruism Community Building*. https://web.archive.org/web/20200220215446/https://www.effectivealtruism.org/groups-effective-altruism-community-building/

Centre for Effective Altruism. (2021a). Payout Report: May-August [November] 2021 - EA Infrastructure Fund Grants. *Effective Altruism Funds*. https://archive.fo/pw6Om

Centre for Effective Altruism. (2021b). *Trustees' Report and Financial Statements for the Year Ended 30 June 2020*. Charity Commission for England and Wales.

Centre for Effective Altruism. (n.d.). *Community Health*. https://www.centreforeffectivealtruism.org/community-health

Christiano, P. (2010, November 27). Comment on: Tallinn-Evans $125,000 Singularity Challenge. *LessWrong*. https://www.lesswrong.com/posts/aqyLxWzAEpDHm2Xyf/tallinn-evans-usd125-000-singularity-challenge?commentId=mZNtZCE4b796x5tQA

Christiano, P. (2013). Influencing the Far Future. *80,000 Hours* https://web.archive.org/web/20140906224435/https://80000hours.org/2013/10/influencing-the-far-future/

Colbourn, G. (2022). Reply to comment on: Free-spending EA might be a big problem... *EA Forum*. https://web.archive.org/web/20221016094945/https://forum.effectivealtruism.org/posts/HWaH8tNdsgEwNZu8B/free-spending-ea-might-be-a-big-problem-for-optics-and?commentId=aRib7T6vWtdj2fi7z

Companies House - GOV.UK. (2012). Certificate of Incorporation of a Private Limited Company - Company No. 7962181 - Centre for Effective Altruism.

Companies House - GOV.UK. (2022a). Certificate of Passing of Centre for Effective Altruism (the "Company")/Certificate of Incorporation on Change of Name. https://tinyurl.com/4pt57tp5

Companies House - GOV.UK. (2022b). Notice of change of name by resolution (Existing Name: Centre for Effective Altruism, Proposed Name: Effective Ventures Foundation). https://find-and-update.

company-information.service.gov.uk/company/07962181/filing-history

Cotton-Barratt, O. (2022). Reply to question: Why did CEA buy Wytham Abbey? [archived 12/7/2022]. *EA Forum*. https://web.archive.org/web/20221207141159/https://forum.effectivealtruism.org/posts/xof7iFB3uh8Kc53bG/why-did-cea-buy-wytham-abbey?commentId=u3yJfbm2pes8TFpYX

Cremer, C. Z. (2020). Objections to Value-Alignment between Effective Altruists. *EA Forum*. https://forum.effectivealtruism.org/posts/DxfpGi9hwvwLCf5iQ/objections-to-value-alignment-between-effective-altruists

Crouch [MacAskill], W. (2010). Hello all. *Felicifia*. https://felicifia.github.io/thread/321.html ; archive at https://web.archive.org/web/20110914020300/http://felicifia.org/viewtopic.php?p=2262#p2248

Crouch [MacAskill], W. (2011, October 22). Banking as an Ethical Career. *Practical Ethics*. http://blog.practicalethics.ox.ac.uk/2011/11/banking-as-an-ethical-career/

Crouch [MacAskill], W. (2012a, February 26). Giving What We Can and 80,000 Hours are recruiting! *Felicifia*. https://web.archive.org/web/20201002005834/https://felicifia.github.io/thread/572.html

Crouch [MacAskill], W. (2012b). Giving What We Can, 80,000 Hours, and Meta-Charity [archived 11/30/2012]. *LessWrong*. https://web.archive.org/web/20121130091244/http://lesswrong.com/lw/fej/giving_what_we_can_80000_hours_and_metacharity/

Crouch [MacAskill], W. (2012c, June 4). Re: Giving to the Centre for Effective Altruism. *Felicifia*. https://felicifia.github.io/thread/614.html#p5507

Crouch [MacAskill], W. (2012d, November 10). Reply to Gedusa on: Giving What We Can, 80,000 Hours and metacharity [archived 11/30/2012]. *LessWrong*. https://web.archive.org/web/20121130091244/http://lesswrong.com/lw/fej/giving_what_we_can_80000_hours_and_metacharity/#:~:text=80k%20members%20give,done%20so%20far.

Crouch [MacAskill], W. (2012e). Responses to questions on donating to 80k, GWWC, EAA and LYCS. *LessWrong*. https://web.archive.org/web/20121130090101/http://lesswrong.com/r/discussion/lw/fkz/responses_to_questions_on_donating_to_80k_gwwc/; current live link https://www.lesswrong.com/posts/rC2EBbhmCzJCSyysN/responses-to-questions-on-donating-to-80k-gwwc-eaa-and-lycs

Crouch [MacAskill], W., & Todd, B. (2011). *Doctor, NGO Worker, or Something Else Entirely? Which Careers Do the Most Good?* https://youtu.be/lAnh2FApskM

Dalton, M. (2018a). I'm pleased to announce the second edition of the effective altruism handbook. *EA Public Facebook group*. https://archive.is/ipPHF , https://archive.is/mguIc

Dalton, M. (2018b, November 8). Why the EA Forum? *EA Forum*. https://forum.effectivealtruism.org/posts/wrPacgwp3DsSJYbce/why-the-ea-forum-1

Dalton, M. (2021). CEA grew a lot in the past year. *EA Forum*. https://forum.effectivealtruism.org/posts/LoSojczPKA5kMchFA/cea-grew-a-lot-in-the-past-year

Daniel, M. (2019). Comment (1) on: Andreas Mogensen's "Maximal Cluelessness". *EA Forum*. https://archive.is/5M6SO#selection-281.0-291.140

Davey, T. (2018). Teaching Today's AI Students To Be Tomorrow's Ethical Leaders: An Interview With Yan Zhang. *Future of Life Institute*. https://futureoflife.org/2018/05/21/teaching-todays-ai-students-to-be-tomorrows-ethical-leaders-yan-zhang/

Deere, S. (2016). Four Ideas You Already Agree With (That Mean You're Probably On Board With Effective Altruism). *Giving What We Can*. https://www.givingwhatwecan.org/post/2016/03/four-things-you-already-agree-with-effective-altruism/

EA Forum. (2022). *Coordination Forum (form. Leaders Forum)*. https://forum.effectivealtruism.org/topics/coordination-forum

EA Hub. (2019a). Tips to Help Your Conversation Go Well. *EA Hub - Resources*. https://archive.md/0EmQK

EA Hub. (2019b). *Tips to Help Your Conversation Go Well: Preventing "Bait and Switch" [archived 10/21/2021]*. https://archive.md/0EmQK#selection-2011.0-2011.28

EA Hub. (2019c). What to Say. *EA Hub - Resources*. https://resources.eahub.org/learn/articles/what-to-say/ ; https://archive.md/Wjtxf

Effective Altruism Funds. (2023). *Grants Database (beta) [archived 3/11/2023]*. https://web.archive.org/

web/20230311103339/https://funds.effectivealtruism.org/grants?sort=round

Effective Altruism Wiki. (2015). Talking About Effective Altruism [archived 12/27/2016]. https://web.archive.org/web/20161227205217/https://effective-altruism.wikia.com/wiki/Talking_about_effective_altruism

Effective Thesis. (2021). Effective Thesis - Home. https://web.archive.org/web/20211215021431/https://effectivethesis.org/

Effective Ventures. (2022). *Effective Ventures - Home [archived 9/23/2022]*. https://web.archive.org/web/20220923212814/https://ev.org/

EffectiveAltruism.org. (2016). *People - Nick Bostrom [archived 9/8/2016]*. https://web.archive.org/web/20160908021556/https://www.effectivealtruism.org/people/nick-bostrom/

EffectiveAltruism.org. (2019). *A Model of an EA Group [archived 2/20/2020]*. https://web.archive.org/web/20200220223435/https://www.effectivealtruism.org/groups-a-model-of-an-ea-group/

Effectivealtruism.org. (2022). *What is Effective Altruism? [archived 10/11/2022]*. https://web.archive.org/web/20221011034320/https://www.effectivealtruism.org/articles/introduction-to-effective-altruism

Elmore, H. (2016). We are in triage every second of every day. *Holly Elmore*. https://mhollyelmoreblog.wordpress.com/2016/08/26/we-are-in-triage-every-second-of-every-day/

Eon Essay Contest. (2022a). *Details*. https://web.archive.org/web/20220417175626/https://www.eonessaycontest.com/details

Eon Essay Contest. (2022b). *Eon Essay Contest on The Precipice*. https://web.archive.org/web/20220224165935/https://www.eonessaycontest.com/

ESPR. (2020). *European Summer Program on Rationality [archived 2/20/2020]*. https://web.archive.org/web/20200220215251/https://espr-camp.org/

European Summer Program in Rationality. (2022). *Staff 2022 [archived 7/18/2022]*. https://web.archive.org/web/20220718133346/https://espr-camp.org/team

Existential Risk Observatory. (2021, January 16). We think writing op-eds about existential risks in mainstream media... *Effective Altruism public Facebook group*. https://archive.is/gqtOo

Existential Risk Observatory. (2022). *Existential Risk Observatory - Home*. https://www.existentialriskobservatory.org/

Flidr, A., & Aung, J. (2018). Heuristics from Running Harvard and Oxford EA Groups. *EA Forum*. https://forum.effectivealtruism.org/posts/Rjb9oTjQ9RJMFw2Yo/heuristics-from-running-harvard-and-oxford-ea-groups

Freeman, L. (2022). "Big tent" effective altruism is very important (particularly right now). *EA Forum*. https://forum.effectivealtruism.org/posts/SjK9mzSkWQttykKu6/big-tent-effective-altruism-is-very-important-particularly

Freeman, L., Basarkod, G., & Wilks, M. (2021). Guide to talking about effective altruism and effective giving. *EA Forum*. https://forum.effectivealtruism.org/posts/qWL7e2uRKJ4CSRT4G/guide-to-talking-about-effective-altruism-and-effective

Freeman, L., Parikh, C., & Giving What we Can. (2022). EA Funds donation platform is moving to Giving What We Can. *EA Forum*. https://web.archive.org/web/20220531172602/https://forum.effectivealtruism.org/posts/3G9w9HfDusFw9LqxQ/ea-funds-donation-platform-is-moving-to-giving-what-we-can

Freeman, L., Townsend, M., Adams, G., Kuhn, F., Hazell, J., & Hartong, J. (2022). GWWC has ambitious plans from 2022 onwards (we're hiring!). *EA Forum*. https://web.archive.org/web/20220516224029/https://forum.effectivealtruism.org/posts/yp4uwxEwamc5LkDAB/gwwc-has-ambitious-plans-from-2022-onwards-we-re-hiring

Fritz, D. (2021). Comment on: Public reports are now optional for EA Funds grantees. *EA Forum*. https://web.archive.org/web/20211203171942/https://forum.effectivealtruism.org/posts/LKdtHdETxSYAXwoW6/public-reports-are-now-optional-for-ea-funds-grantees?commentId=Q5CacKH3qiTDKSJGN

FTX Foundation. (2021). FTX EA Fellowships. *EA Forum*. https://forum.effectivealtruism.org/posts/sdjcH7KAxgB328RAb/ftx-ea-fellowships

Future of Humanity Institute. (2007). *Staff [archived 2/19/2007]*. https://web.archive.org/web/20070219012923/http://www.fhi.ox.ac.uk/staff.html

Future of Humanity Institute. (2011, November). Prize Competition: Crucial Considerations for the Future of Humanity [archived 11/10/2011]. *Future of Humanity Institute*. https://web.archive.org/web/20111110154333/http://www.fhi.ox.ac.uk/prize

Future of Humanity institute. (2012). Prize Competition: Crucial Considerations for the Future of Humanity - Winning entries. *Future of Humanity Institute*. https://web.archive.org/web/20221031151935/https://www.fhi.ox.ac.uk/crucial-considerations-for-the-future-of-humanity-competition-results/ ;old archive, inactive outlinks: https://web.archive.org/web/20120301180305/http://www.fhi.ox.ac.uk/prize

Garth, A. (2022). EA is becoming increasingly inaccessible, at the worst possible time. *EA Forum*. https://forum.effectivealtruism.org/posts/duPDKhtXTJNAJBaSf/ea-is-becoming-increasingly-inaccessible-at-the-worst

Gertler, A. (2017). Talking About Effective Altruism at Parties. *EA Forum*. https://forum.effectivealtruism.org/posts/vwDjfnJ9656fAQvsC/talking-about-effective-altruism-at-parties

Gertler, A. (2019, November 13). EA Leaders Forum: Survey on EA Priorities (data and analysis). *EA Forum*. https://forum.effectivealtruism.org/posts/TpoeJ9A2G5Sipxfit/ea-leaders-forum-survey-on-ea-priorities-data-and-analysis

Gertler, A. (n.d.). Talking About EA. https://docs.google.com/document/d/1Q9P9HXN4rzuk3Fhg12hmVEwanSxfdBZZxUDL-_UG5FE/edit

GiveWell. (2007a). *The Case for the Clear Fund*. GiveWell. https://www.givewell.org/wiki/index.php?title=Clear_Fund_Business_Plan

GiveWell. (2007b, May 5). *Guide to GiveWell*. Retrieved May 22 from https://web.archive.org/web/20070702193028/http://www.givewell.net/wiki/index.php?title=Guide_to_GiveWell

GiveWell. (2007c). *Welcome to GiveWell [archived 2/9/2007]*. https://web.archive.org/web/20070209143806/http://www.givewell.net/wiki/index.php?title=Welcome_to_GiveWell

GiveWell. (2013a). *Clear Fund Board Meeting - January 24, 2013*. https://web.archive.org/web/20130430150803/http://www.givewell.org/about/official-records/board-meeting-21

GiveWell. (2013b). *Mass Distribution of Long-Lasting Insecticide-Treated Nets (LLINs)*. https://www.givewell.org/international/technical/programs/insecticide-treated-nets

Giving What We Can. (2007). *Giving What We Can*. Retrieved July 1 from http://web.archive.org/web/20070701205101/http://www.givingwhatwecan.org/

Giving What We Can. (2011). *Our Pledge [archived 8/13/2011]*. https://web.archive.org/web/20110813205139/http://www.givingwhatwecan.org/our-pledge/

Giving What We Can. (2012). *Starter Packet For Giving What We Can Chapters (aka "Big Awesome Super Duper Starter Packet")*. https://docs.google.com/document/d/1mRDbZgRIefzQ45w6T8xFuluFBFkIc0WNnUX5dpKqjuY/edit

Giving What We Can. (2016). *Workplace Activism: EA Handbook*. https://web.archive.org/web/20171118113912/http://www.eaworkplaceactivism.org/wp-content/uploads/2016/10/WA-Handbook-v1.pdf

Giving What We Can. (n.d.). *Effective Altruism: Chapter Starter Packet*. https://drive.google.com/file/d/0BxN9jkWGI8LFdEFLS0VfNTBtbkE/view

Global Challenges Project. (2021a). *EA Student Groups Handbook (Packaged Programs)*. https://handbook.globalchallengesproject.org/

Global Challenges Project. (2021b). *Packaged Program - Intro Talks: our ready-to-go EA intro talk transcript*. https://web.archive.org/web/20211214113912/https://handbook.globalchallengesproject.org/packaged-programs/intro-talks

Global Challenges Project. (2022a). *Global Challenges Project - Home [archived 11/8/2022]*. https://web.archive.org/web/20221108202736/https://www.globalchallengesproject.org/

Global Challenges Project. (2022b). Reframing Student Groups. https://web.archive.org/web/20220704000340/https://handbook.globalchallengesproject.org/guides/reframing-student-groups

Habryka, O. (2021). Comment on: Towards a Weaker Longtermism. *EA Forum*. https://forum.effectivealtruism.org/posts/fStCX6RXmgxkTBe73/towards-a-weaker-longtermism?commentId=widZvbTdt88PF9GaY

Habryka, O. (2022a). Comment on: Will Lightcone purchase the Rose Garden Inn? *Manifold Markets*. https://manifold.markets/noumena/will-lightcone-purchase-the-rose-ga#ERP9oXnuKyrUgToupHQ7; https://archive.vn/3zUYV#selection-553.0-553.81

Habryka, O. (2022b, December 10). Reply to comment on: Why did CEA buy Wytham Abbey? *EA Forum*. https://web.archive.org/web/20221212092829/https://forum.effectivealtruism.org/posts/xof7iFB3uh8Kc53bG/why-did-cea-buy-wytham-abbey?commentId=tSEwK79JGJJuEkhTe

Hazell, J. (2022a, September 21). A more convincing case for longtermism. *Muddy Clothes*. https://muddyclothes.substack.com/p/a-more-convincing-case-for-longtermism

Hazell, J. (2022b, July 31). Why I view effective giving as complementary to direct work. *EA Forum*. https://forum.effectivealtruism.org/posts/QFP4xdcrmPSyPprr6/why-i-view-effective-giving-as-complementary-to-direct-work

HM Land Registry -GOV.UK. (2022). Title Register for: 118 Abingdon Road, Oxford, OX1 4PZ (Freehold), Title Number: ON83507.

HM Land Registry -GOV.UK. (2023). Title register for: Wytham Abbey, Wytham, Oxford, OX2 8QB (Freehold), Title number: ON211059.

Holness-Tofts, A. (2021). Comment on: Towards a Weaker Longtermism. *EA Forum*. https://forum.effectivealtruism.org/posts/fStCX6RXmgxkTBe73/towards-a-weaker-longtermism?commentId=b4S2jhgYcNzCXgo23

Howe, M. (2002, June 24). A Place to Stand. *Accelerating Future*. https://web.archive.org/web/20040803100942/http://www.acceleratingfuture.com/articles/aplacetostand.htm

Hurford, P. (2012, July 19). Giving Is Hard, But There's Help. *Great Play*. https://web.archive.org/web/20121022042309/http://www.greatplay.net/essays/giving-is-hard-but-theres-help

Hurford, P. (2014, May 18). Conversation with Michael Bitton about EA Marketing. *Everyday Utilitarian*. https://web.archive.org/web/20141030073244/http://everydayutilitarian.com/essays/conversation-with-michael-bitton-about-ea-marketing/

Hurford, P., & Barnett, T. (2017). EA Survey 2017 Series: Have EA Priorities Changed Over Time? *EA Forum*. https://forum.effectivealtruism.org/posts/A54Lrz5vQK5kgtxA2/ea-survey-2017-series-have-ea-priorities-changed-over-time

Hutchinson, M. (2016). Why Poverty? *EA Forum*. https://web.archive.org/web/20220828155524/https://forum.effectivealtruism.org/posts/h2kGhaTYtnd7M8fov/why-poverty

Illing, S., & MacAskill, W. (2018, March 2). How to do good better: The case for effective altruism, according to one of its founders. *Vox - Future Perfect*. https://www.vox.com/future-perfect/2018/12/14/18088514/effective-altruism-poverty-philanthropy-william-macaskill

Intelligence Squared. (2015). *Effective Altruism: A Better Way to Lead an Ethical Life (Debate between Will MacAskill and Giles Fraser moderated by Kamal Ahmed)* https://www.intelligencesquared.com/events/effective-altruism-a-better-way-to-lead-an-ethical-life/

Jacobs, A. J. (2019). The Billionaire and the (Very, Very) Far Future (interview with Natalie Cargill and Ben Delo). https://www.linkedin.com/pulse/billionaire-very-far-future-a-j-jacobs/

Jacobs, B. (2022). ESPR should return the FTX-funded chateau. *EA Forum*. https://forum.effectivealtruism.org/posts/pbe8x4AQDqftQoaT5/espr-should-return-the-ftx-funded-chateau

Jeyapragasan, K. (2021). How students, groups, and community members can use funding. *EA Forum*. https://forum.effectivealtruism.org/posts/hHtyHW368b3jjHLHM/how-students-groups-and-community-members-can-use-funding

Jonathan Tuckey Design. (2022). *Wytham Abbey: A New Retreat for Discussion*. https://jonathantuckey.com/wytham-abbey-a-new-retreat-for-discussion/

Kadosh, G. (2022). How EA is perceived is crucial to its future trajectory. *EA Forum*. https://forum.effectivealtruism.org/posts/82ig8odF9ooccfJfa/how-ea-is-perceived-is-crucial-to-its-future-trajectory

Karnofsky, H. (2007). Here's to fear of failure. *GiveWell blog*. https://blog.givewell.org/2007/11/20/heres-to-fear-of-failure/

Karnofsky, H. (2011). Announcing GiveWell Labs. *GiveWell blog*. https://web.archive.org/web/20120124235852/http://blog.givewell.org/2011/09/08/announcing-givewell-labs/

Karnofsky, H. (2016a). Hits-based Giving. *Open Philanthropy Project*. https://www.openphilanthropy.org/blog/hits-based-giving

Karnofsky, H. (2016b). Our Progress in 2015 and Plans for 2016. *Open Philanthropy Project*. https://www.openphilanthropy.org/research/our-progress-in-2015-and-plans-for-2016/

Karnofsky, H. (2016c, May 16). Potential Risks from Advanced Artificial Intelligence: The Philanthropic Opportunity. https://web.archive.org/web/20190623073146/https:/www.openphilanthropy.org/blog/potential-risks-advanced-artificial-intelligence-philanthropic-opportunity

Karnofsky, H. (2016d). Some background on our views regarding advanced artificial intelligence. *Open Philanthropy Project blog*. http://www.openphilanthropy.org/blog/some-background-our-views-regarding-advanced-artificial-intelligence

Karnofsky, H. (2016e, March 6). Three Key Issues I've Changed My Mind About. *Open Philanthropy Project*. https://web.archive.org/web/20161011153556/https://www.openphilanthropy.org/blog/three-key-issues-ive-changed-my-mind-about

Karnofsky, H. (2017). The Open Philanthropy Project is Now an Independent Organization. *Open Philanthropy Project*. https://www.openphilanthropy.org/blog/open-philanthropy-project-now-independent-organization#LLC

Karnofsky, H. (2021a). The "most important century" blog post series. *Cold Takes*. https://archive.is/GPwbx

Karnofsky, H. (2021b). Roadmap for the "most important century" series. *Cold Takes*. https://archive.is/bFIZH

Kaufman, J. (2012, September 18). History of "Earning to Give". *jefftk*. https://www.jefftk.com/p/history-of-earning-to-give

Knutsson, S. (2019). Problems in effective altruism and existential risk and what to do about them. *Simon Knutsson*. https://www.simonknutsson.com/problems-in-effective-altruism-and-existential-risk-and-what-to-do-about-them/

Koehler, A., Todd, B., Wiblin, R., & Harris, K. (2020). Benjamin Todd on the core of effective altruism and how to argue for it. *80,000 Hours podcast*. https://80000hours.org/podcast/episodes/ben-todd-on-the-core-of-effective-altruism/

Kotikova, I. (2022). Reply to comment by Jan Kulveit on: Why did CEA buy Wytham Abbey? *EA Forum*. https://forum.effectivealtruism.org/posts/xof7iFB3uh8Kc53bG/why-did-cea-buy-wytham-abbey?commentId=Bjf7TiPumKAcRSzi6

Kotíková, I. (2022). Announcing: the Prague Fall Season. *EA Forum*. https://web.archive.org/web/20220729190814/https://forum.effectivealtruism.org/posts/ZxHnaPmxmBrZmerBj/announcing-the-prague-fall-season

Kulveit, J. (2018). Comment on: Max Dalton: I'm pleased to announce the second edition of the effective altruism handbook. https://web.archive.org/web/20200929190922/https://www.facebook.com/groups/effective.altruists/permalink/1750780338311649/?comment_id=1752865068103176

Kulveit, J. (2022). Reply to: Why did CEA buy Wytham Abbey? *EA Forum*. https://archive.vn/t5kEe

LEAF. (2021a). '*Building a Better Future*' *Course [archived 12/25/2021]*. https://web.archive.org/web/20211225104505/https://www.leaf.courses/

LEAF. (2021b). *Home - Leaf [archived 10/06/2021]*. https://web.archive.org/web/20211006072438/https://www.leaf.courses/home

LEAF. (2021c). *Leaf - Minicourse [archived 12/25/2021]*. https://archive.ph/UFluv

Leong, C. (2017). Comment on: The hidden cost of shifting away from poverty. *EA Forum*. https://forum.effectivealtruism.org/posts/MrCjtdTxHvRzqu8fp/the-hidden-cost-of-shifting-away-from-poverty?commentId=TbZ4645kw2gNWM43w

Leong, C. (2022a). Comment on: Making Effective Altruism Enormous. *EA Forum*. https://forum.effectivealtruism.org/posts/cvzEpGgmQ8BefiwGt/making-effective-altruism-enormous?commentId=RdTuakPQQZJKcfivY

Leong, C. (2022b). Should EA be explicitly long-termist or uncommitted? *EA Forum*. https://forum.

effectivealtruism.org/posts/cnvejLCdG9bpPxLNg/should-ea-be-explicitly-long-termist-or-uncommitted

Lewis, G. (2016). Beware surprising and suspicious convergence. https://gregoryjlewis.com/2016/01/24/beware-surprising-and-suspicious-convergence/

Lewis, G. (2018). Comment on: Announcing the Effective Altruism handbook, second edition. *EA Forum*. https://web.archive.org/web/20220812124030/https://forum.effectivealtruism.org/posts/MQWAsdm8MSzYNkD9X/announcing-the-effective-altruism-handbook-2nd-edition?commentId=BSzBCdYYfFcCKoP5P

Lifland, E. (2022). Prioritizing x-risks may require caring about future people. *EA Forum*. https://forum.effectivealtruism.org/posts/rvvwCcixmEep4RSjg/prioritizing-x-risks-may-require-caring-about-future-people

Longview Philanthropy, & Effective Ideas. (2022a). The Blog Prize [archived 4/4/2022]. *Effective Ideas*. https://web.archive.org/web/20220308132010/https://effectiveideas.org/

Longview Philanthropy, & Effective Ideas. (2022b). Effective Ideas is Announcing $500,000 Blog Prize [archived: 4/4/2022]. *Effective Ideas*. https://web.archive.org/web/20220404164400/https://effectiveideas.org/effective-ideas-is-announcing-500000-blog-prize/

Longview Philanthropy [formerly Effective Giving UK]. (2020). *Longview Philanthropy - Grantmaking*. Retrieved July 20 from https://web.archive.org/web/20210204221457/https://www.longview.org/grantmaking

MacAskill, A., & MacAskill, W. (2015, September 9). To truly end animal suffering, the most ethical choice is to kill wild predators (especially Cecil the lion). *Quartz*. https://qz.com/497675/to-truly-end-animal-suffering-the-most-ethical-choice-is-to-kill-all-predators-especially-cecil-the-lion/

MacAskill, W. (2013a, October 2). It's OK to have a big carbon footprint and pay others not to. *Quartz*. https://qz.com/131024/its-ok-to-have-a-big-carbon-footprint-and-pay-others-not-to/

MacAskill, W. (2013b, February 27). To save the world, don't get a job at a charity; go work on Wall Street. *Quartz*. https://qz.com/57254/to-save-the-world-dont-get-a-job-at-a-charity-go-work-on-wall-street/

MacAskill, W. (2013c, December 3). What Charity Navigator Gets Wrong About Effective Altruism. *Stanford Social Innovation Review*. https://ssir.org/articles/entry/what_charity_navigator_gets_wrong_about_effective_altruism

MacAskill, W. (2013d). What Warren Buffett's son doesn't understand about the world. *Quartz*. https://qz.com/110616/what-peter-buffett-doesnt-understand-about-the-world/

MacAskill, W. (2013e, May 16). What we can learn from one of the worst charities in the world. *Quartz*. https://qz.com/84943/what-we-can-learn-from-one-of-the-worst-charities-in-the-world/

MacAskill, W. (2013f). Which field of law should you go into? *Petrie Flom blog - Harvard Law*. https://blog.petrieflom.law.harvard.edu/2013/06/04/which-field-of-law-should-you-go-into/

MacAskill, W. (2014a, August 14). The cold, hard truth about the ice bucket challenge. *Quartz*. https://qz.com/249649/the-cold-hard-truth-about-the-ice-bucket-challenge/

MacAskill, W. (2014b, April 30). Peter Buffett: Philanthropy disguises itself as a fix when it is part of the problem. *Quartz*. https://qz.com/201778/warren-buffetts-son-philanthropy-disguises-itself-as-a-fix-when-its-a-part-of-the-problem/

MacAskill, W. (2014c). Replaceability, Career Choice, and Making a Difference. *Ethical Theory and Moral Practice*, *17*(2), 269-283.

MacAskill, W. (2014d, August 18). This week, let's dump a few ice buckets to wipe out malaria too. *Quartz*. https://qz.com/250845/this-week-lets-dump-a-few-ice-buckets-to-wipe-out-malaria-too/

MacAskill, W. (2015a, December 4). 5 Criticisms of billionaire mega-philanthropy debunked. *Quartz*. https://qz.com/564805/5-criticisms-of-billionaire-mega-philanthropy-debunked/

MacAskill, W. (2015b). *Doing Good Better: How Effective Altruism Can Help You Make a Difference*. Gotham Books.

MacAskill, W. (2015c, December 2). Mark Zuckerberg is the perfect model for how anyone can change the world. *Quartz*. https://qz.com/563325/mark-zuckerbergs-donation-is-a-perfect-model-for-how-anyone-can-change-the-world/

MacAskill, W. (2015d, October 9). This Couple Lives On 6% Of Their Income So They Can Give $100,000 A Year To Charity. *Quartz*. https://qz.com/515655/this-couple-lives-on-6-of-their-income-so-they-can-give-100000-a-year-to-charity/

MacAskill, W. (2015e). Want to make a difference? Don't work for a charity. *TEDx Cambridge University*. https://youtu.be/hU5O4PhrOik

MacAskill, W. (2017, March 29). Doing good, and doing it better. https://portland-communications.com/development-global-issues/doing-good-and-doing-it-better/

MacAskill, W. (2019). The Definition of Effective Altruism. In H. Greaves & T. Pummer (Eds.), *Effective Altruism: Philosophical Issues* (pp. 10-28). Oxfod University Press.

MacAskill, W. (2022a, August 5). The Case for Longtermism. *The New York Times*. https://www.nytimes.com/2022/08/05/opinion/the-case-for-longtermism.html

MacAskill, W. (2022b). Comment on: "Longtermism" vs. "Existential Risk". *EA Forum*. https://web.archive.org/web/20220817121747/https://forum.effectivealtruism.org/posts/KDjEogAqWNTdddF9g/long-termism-vs-existential-risk?commentId=eXpqe8mHEEHTLy9jL

MacAskill, W. (2022c). EA and the Current Funding Situation. *EA Forum*. https://forum.effectivealtruism.org/posts/cfdnJ3sDbCSkShiSZ/ea-and-the-current-funding-situation

MacAskill, W. [@willmacaskill]. (2022d). *I didn't always buy these ideas [note: he is referring to 'longtermism' and the far future]. For much of my life, I spent my energy on more visceral problems - my first book, Doing Good Better, was all about how we can most effectively improve the lives of the extreme poor, such as by funding anti-malarial bednets*. Twitter. https://twitter.com/willmacaskill/status/1554378997609353216

MacAskill, W. (2022e). *What We Owe the Future*. Basic Books.

MacAskill, W. (n.d.). *The definition of effective altruism*. https://drive.google.com/file/d/1rQu75k8uMFpdsp1y3JWlHP6kev3T-97N/view

Malagon, S. (2022). Mexico EA Fellowship. *EA Forum*. https://forum.effectivealtruism.org/posts/kBgMD7GwJboxGFvsp/mexico-ea-fellowship

Matheny, J. G. (2006). Reducing the risk of human extinction. *Accelerating Future*. https://web.archive.org/web/20061024143031/http://www.acceleratingfuture.com/papers/extinction.htm

Matheny, J. G. (2009). *Ought we worry about human extinction?* https://web.archive.org/web/20090403104720/http:/jgmatheny.org/extinctionethics.htm

Matthews, D. (2016, March 8). For the love of God, rich people, stop giving Stanford money. *Vox*. https://www.vox.com/2016/2/26/11117200/stanford-phil-knight-hennessy-scholarship

Mayhew, S. (2019, September 8). What to know before talking to journalists about EA. *EA Forum*. https://forum.effectivealtruism.org/posts/6hbyTXT5kvDmGp3AY/what-to-know-before-talking-with-journalists-about-ea

Melchin, D. (2020). Comment on: New data suggests the 'leaders" priorities represent the core of the community. *EA Forum*. https://forum.effectivealtruism.org/posts/DoAWvhgEnhSMr6yjS/new-data-suggests-the-leaders-priorities-represent-the-core?commentId=yavATsqfatMgDwwZv

Michaud, P. (2017, February 24). CFAR 2017 Venue Update. *Center for Applied Rationality* https://www.rationality.org/resources/updates/2017/cfar-2017-venue-update

Mindermann, S. (2013). My Introduction to Felicifia. *Felicifia*. https://felicifia.github.io/thread/933.html

Moorhouse, F. (2022). Introduction to Longtermism [archived 3/2/2022]. *EffectiveAltruism.org*. https://web.archive.org/web/20220302173251/https://www.effectivealtruism.org/articles/longtermism

Nachbauer, T. (2011). *Transhuman - Do you want to live forever?* https://www.youtube.com/watch?v=3PAj2yorJig

Nanda, N. (2022). Simplify EA Pitches to "Holy Shit, X-Risk". *AI Alignment Forum*. https://www.alignmentforum.org/posts/aCzJKKFN3RC3zsJSD/simplify-ea-pitches-to-holy-shit-x-risk

Nonlinear. (2022). *The Productivity Fund: A low-barrier fund paying for productivity enhancing tools for longtermists*. https://www.nonlinear.org/productivity-fund.html ; archive https://web.archive.org/web/20220803123336/https://www.nonlinear.org/productivity-fund.html

O'Keefe, C. (2018). A *Vision for* HUEASG 2018-19. https://docs.google.com/document/d/1i1-57jRg7vrcTBXAcqIFAYzGoC-bftIMnElqPKFGMVk/edit?usp=sharing

Open Philanthropy. (2017). Grant - Center for Applied Rationality - European Summer Program on Rationality (ESPR) ($340,000). https://www.openphilanthropy.org/grants/center-for-applied-rationality-european-summer-program-on-rationality/

Open Philanthropy. (2018). *Grant - Course development support for David Manley for "Changing the World" at University of Michigan* ($46,696). https://web.archive.org/web/20181128191539/https://www.openphilanthropy.org/giving/grants/university-of-michigan-support-for-david-manley

Open Philanthropy. (2019). *Grant - European Summer Program on Rationality* (ESPR), *General Support* ($510,000). https://www.openphilanthropy.org/grants/european-summer-program-on-rationality-general-support/

Open Philanthropy. (2021a). *Grant - Course development support: Jacob Steinhardt, Danny Hernandez, Yan Zhang at UC Berkeley, course on forecasting* ($53,983). https://web.archive.org/web/20220929071751/https://www.openphilanthropy.org/grants/uc-berkeley-course-development-support-jacob-steinhardt/

Open Philanthropy. (2021b). Grant - Course development support: Michael Specter co-teaching with Kevin Esvelt ($85,000). https://www.openphilanthropy.org/grants/michael-specter-course-development-support/

Open Philanthropy. (2021c). *Grant - Course development support: University of Southern California - Jake Nebel* ($33,000). https://www.openphilanthropy.org/grants/university-of-southern-california-course-development-support-jake-nebel/

Open Philanthropy. (2021d). *Open Philanthropy Course Development Grants*. https://web.archive.org/web/20211028090235/https://www.openphilanthropy.org/focus/other-areas/open-philanthropy-course-development-grants ;

Open Philanthropy. (2021* [2023]). *Grant - Effective Ventures Foundation UK, Event Venue* ($22,805,823) [*note: the grant award report backdates the award date to December 2021, however, the grant was not included in Open Philanthropy's grants database until February 2023]. https://web.archive.org/web/20230220161719/https://www.openphilanthropy.org/grants/effective-ventures-foundation-uk-event-venue/

Open Philanthropy. (2022a). *Grant - CEA/Harvard AI-Safety Team (HAIST) office space* ($250,000). https://web.archive.org/web/20221106100558/https://www.openphilanthropy.org/grants/centre-for-effective-altruism-harvard-ai-safety-office/

Open Philanthropy. (2022b). *Grant - Centre for Effective Altruism, Biosecurity Coworking Space* ($5,318,000). https://web.archive.org/web/20221216132005/https://www.openphilanthropy.org/grants/centre-for-effective-altruism-biosecurity-coworking-space/

Open Philanthropy. (2022c). *Grant - Centre for Effective Altruism, Harvard Square Coworking Space* ($8,875,000). https://web.archive.org/web/20230129184304/https://www.openphilanthropy.org/grants/centre-for-effective-altruism-harvard-coworking-space/

Open Philanthropy. (2022d). *Grant - Course development support, Matti Wilks, University of Edinburgh* ($47,000). https://web.archive.org/web/20220929112107/https://www.openphilanthropy.org/grants/university-of-edinburgh-effective-altruism-psychology-course/

Open Philanthropy. (2022e). *Grant - Course development support, William D'Alessandro at Ludwig Maxilian U. of Munich*, $19,000. https://web.archive.org/web/20220929064837/https://www.openphilanthropy.org/grants/william-dalessandro-course-development-support/

Open Philanthropy. (2022f). *Grant - Course development support: University of Michigan, Anna Edmonds* ($50,300). https://web.archive.org/web/20220708075444/https://www.openphilanthropy.org/grants/university-of-michigan-course-development-support-anna-edmonds/

Open Philanthropy. (2022g). *Grant - Course development support: Hugo Viciana and Aníbal Monasterio Astobiza, Universidad de Sevilla* ($19,982). https://web.archive.org/web/20220929104844/https://www.openphilanthropy.org/grants/universidad-de-sevilla-course-on-effective-altruism/

Open Philanthropy. (2022h). *Grant - Course Development Support: Princeton University Existential Risk Course Development* ($27,555) *for Adam Elga and Kevin Baker*. https://web.archive.org/web/20221231214846/https://www.openphilanthropy.org/grants/princeton-university-existential-risk-course-development/

Open Philanthropy. (2022i). *Grant - Course Development Support: Seton Hall University/Travis Timmerman philosophy course "Longtermism: Past, Present, and Future"* ($39,780). https://www.openphilanthropy.org/grants/seton-hall-university-longtermism-course-development/

Open Philanthropy. (2022j). *Grant - Course Development Support: Trevor Woolley, Effective Altruism Economics Course* ($10,350). https://web.archive.org/web/20221231214902/https://www.openphilanthropy.org/grants/trevor-woolley-effective-altruism-economics-course/

Open Philanthropy. (2022k). *Grant - European Summer Program on Rationality* (ESPR), *General Support* ($4,715,000). https://www.openphilanthropy.org/grants/european-summer-program-on-rationality-general-support-2022/

Open Philanthropy. (2022l). *Grant - London EA Hub/EA Organizing in London* ($445,000). https://web.archive.org/web/20230220095340/https://www.openphilanthropy.org/grants/london-effective-altruism-hub-effective-altruism-organizing-in-london/

Open Philanthropy. (2022m). *Victory Briefs Institute - Effective Altruism Curriculum for Debaters* ($95,000). https://www.openphilanthropy.org/focus/effective-altruism-community-growth/victory-briefs-ea-curriculum

Ord, T. (2007). Giving What We Can. *Felicifia*. https://web.archive.org/web/20090818015737/http://www.felicifia.com/showDiary.do?diaryId=135

Ord, T. (2011, January 28). First Person: Toby Ord. *Financial Times*. https://www.ft.com/content/25a4ecd6-28ef-11e0-aa18-00144feab49a

Ord, T. (2020). *The Precipice*.

Ough, T. (2017, May 31). Don't follow your passion and think twice about medicine: the new rules of choosing a career. *The Telegraph*. https://www.telegraph.co.uk/education-and-careers/0/dont-follow-passion-think-twice-medicine-new-rules-choosing/

Patel, A. (2022, May 12). A hypothesis for why some people mistake EA for a cult. *EA forum*. https://forum.effectivealtruism.org/posts/TcKbXwBX7YJ4NfgGL/a-hypothesis-for-why-some-people-mistake-ea-for-a-cult

Pearce, A. (2015, August 10). *Response to Rob Wiblin, "Existential-risk focussed people should be more polite..."* https://archive.md/elTRO#selection-1245.0-1245.251

Pieters, J. W., Formanek, C., & Ghoor, A. (2022). Cape Town Summer EA Office. *EA Forum*. https://web.archive.org/web/20220831145433/https://forum.effectivealtruism.org/posts/pXkt7oRczDrqvTu4B/cape-town-summer-ea-office

Piper, K. (2014). The Unit of Caring - home [archived 10/3/2014]. *The Unit of Caring*. https://web.archive.org/web/20141003215310/http://theunitofcaring.tumblr.com/

Piper, K. (2016). The Unit of Caring - About [archived 1/31/2016]. *The Unit of Caring*. https://web.archive.org/web/20160131060710/http://theunitofcaring.tumblr.com/about

Piper, K. (2018). Vox has a new department, Future Perfect, covering the world from an effective altruist perspective. *EA Forum*. https://forum.effectivealtruism.org/posts/areeSd8Z4rQXjjL4y/vox-has-a-new-department-future-perfect-covering-the-world

Plant, M. (2022). Reply to comment by AnonymousEAForumAccount on: EA is more than longtermism (originally: Is EA just longtermism now?). *EA Forum*. https://web.archive.org/web/20220527134709/https://forum.effectivealtruism.org/posts/LRmEezoeeqGhkWm2p/ea-is-more-than-longtermism?commentId=fvgBryEmY6ztyvPqw

Prague Fall Season. (2022). *Prague Fall Season - home* https://web.archive.org/web/20220812142349/https://praguefallseason.com/

Prometheus Science. (2022). *Prometheus Science Bowl [archived 1/12/2022]*. https://web.archive.org/web/20220112044748/https://prometheus.science/

Pummer, T. (2016). Whether and where to give. *Philosophy & Public Affairs*, 44(1), 77–95.

Redwood, Z. (2011, July 26). The best causes - updated. *80,000 Hours*. https://80000hours.org/2011/11/the-best-causes-updated/

Redwood, Z. (2012a, April 12). Living to 1000: an interview with Aubrey de Grey. *80,000 Hours*. https://web.archive.org/web/20141117052116/https://80000hours.org/2012/04/living-to-1000-an-interview-with-aubrey-de-grey/

Redwood, Z. (2012b). The Singularity Institute is seeking remote researchers (writing skill not required). *80,000 Hours*. https://80000hours.org/2012/04/the-singularity-institute-is-seeking-remote-researchers-writing-skill-not-required/

Redwood, Z., & Muehlhauser, L. (2012). High Impact Interview 1: Existential Risk Research at SIAI [archived 5/2/2012]. *80,000 Hours*. https://web.archive.org/web/20120502223553/http://80000hours.org/blog/17-high-impact-interview-1-existential-risk-research-at-siai

Reese, J. (2015). EA Interview Series: Michelle Hutchinson, December 2015. *EA Forum*. https://web.archive.org/web/20160225011320/http://effective-altruism.com/ea/rm/ea_interview_series_michelle_hutchinson_december/

Rethink Charity. (2018). *Local Effective Altruism Network (LEAN) [archived 1/16/2018]*. https://web.archive.org/web/20180116032348/https://rtcharity.org/lean/

Rethink Wellbeing. (2023). *Rethink Wellbeing - Home [archived 2/26/2023]*. https://web.archive.org/web/20230226102317/https://www.rethinkwellbeing.org/

Rohrig, A., Dalton, M., & Ross, N. (2022, August 2). Some updates in EA communications. *EA Forum*. https://web.archive.org/web/20220805065830/https://forum.effectivealtruism.org/posts/g9gj6AtesiTMwNETi/some-updates-in-ea-communications

Rosenfeld, G. (2022). Free-spending EA might be a big problem for optics and epistemics. *EA Forum*. https://forum.effectivealtruism.org/posts/HWaH8tNdsgEwNZu8B/

Sandberg, A. (2015). Transhumanism and the Meaning of Life. In C. Mercer & T. J. Trothen (Eds.), *Religion and Transhumanism: The Unkonw Future of Human Enhancement* (pp. 3-22). Praeger.

Savills. (2021). *The Savills Portfolio: Exclusive Insights into the World's Most Desirable Homes (vol. 4, 2021)*. https://pdf.savills.com/edocs/the-savills-portfolio-2021/102-103/index.html

Savoie, J. (2018, August 3). Problems with EA representativeness and how to solve it. *EA Forum*. https://forum.effectivealtruism.org/posts/2qGL7FoTEZTwp38ij/problems-with-ea-representativeness-and-how-to-solve-it

Scholl, A. (2019). CFAR: Progress Report & Future Plans. https://www.lesswrong.com/posts/vj6CYLuDPw3ieCB4A/cfar-progress-report-and-future-plans-1

Shlegeris, B. (2020). Some thoughts on EA outreach to high schoolers. *EA Forum*. https://forum.effectivealtruism.org/posts/HcaB2kJKhxJtS4oGc/some-thoughts-on-ea-outreach-to-high-schoolers

Shulman, C. (2007, June 4). re: optimal action (comment on Existential Risks). *Felicifia*. http://web.archive.org/web/20070901224824/http://felicifia.com/showDiary.do?diaryId=67

Singer, P. (1972). Famine, Affluence, and Morality. *Philosophy & Public Affairs*, 1(3), 229-243. http://www.jstor.org/stable/2265052

Singer, P. (1997). The Drowning Child and the Expanding Circle. *New Internationalist*. https://newint.org/features/1997/04/05/peter-singer-drowning-child-new-internationalist

Singer, P. (2013). The Why and How of Effective Altruism (transcript). *TED Talks*. https://www.ted.com/talks/peter_singer_the_why_and_how_of_effective_altruism

Singer, P. (2015). *The Most Good You Can Do: How Effective Altruism is Changing Ideas about Living Ethically*. Yale University Press.

Singer, P. (2021). The Hinge of History. *Project Syndicate*. https://www.project-syndicate.org/commentary/ethical-implications-of-focusing-on-extinction-risk-by-peter-singer-2021-10

Singer, P. (n.d.). Course: Effective Altruism. *Coursera*. https://www.coursera.org/learn/altruism

Singer, P., & Wong, B. (2019). Interview with Peter Singer (June 3, 2019). *Oxford Political Review*. https://fb.watch/dg_3ed3C7p/

Singularity Institute for Artificial Intelligence. (2002). *Six Reasons Why Small Donations Matter*. https://web.archive.org/web/20021220084208/http://singinst.org/small-donations-matter.html

Sittler [Adamczewski], T. (2015). Altruism Icebreakers. *The Life You Can Save*. https://www.thelifeyoucansave.org/blog/altruism-icebreakers/

Stafforini, P. (2012a, October 27). Comment on: Astronomically astronomical waste. *Felicifia*. https://web.archive.org/web/20130619045119/http://www.felicifia.org/viewtopic.php?f=29&t=715

Stafforini, P. (2012b, July 16). Comment on: How can I reduce existential risk from AI. *LessWrong*. https://www.lesswrong.com/posts/qARBe3jBodrdPeRE6/how-can-i-reduce-existential-risk-from-ai?commentId=Xd3N7fP6wSJXLDYzz

Stafforini, P. (2012c, July 3). Suggested list of readings for utilitarian landing page. *Felicifia*. https://felicifia.github.io/thread/640.html

Tallinn, J. (2020). Philanthropy [archived: 10/15/2020]. *JaanOnline*. https://web.archive.org/web/20201015093153/https://jaan.online/philanthropy.html

Telleen-Lawton, T. (2019). *CFAR Newsletter (May 2019)*. https://www.rationality.org/resources/updates/2019/may-newsletter

THINK. (2013). *Intro to Effective altruism (THINK Module #1)*. https://web.archive.org/web/20130817095817/http://www.thehighimpactnetwork.org/modules/intro-to-effective-altruism

Todd, B. (2013a). How Important are Future Generations? *80,000 Hours*. https://web.archive.org/web/20130829023754/http://80000hours.org/blog/245-how-important-are-future-generations

Todd, B. (2013b, October 27). The worst ethical careers advice in the world. *80,000 Hours*. https://80000hours.org/2013/05/the-worst-ethical-careers-advice-in-the-world/

Todd, B. (2014). Why long-run focused effective altruism is more common sense. *EA Forum*. https://forum.effectivealtruism.org/posts/WdtcxCKFLLCeLBR6H/why-long-run-focused-effective-altruism-is-more-common-sense

Todd, B. (2017). Reply to comment on: The marketing gap and a plea for moral inclusivity. https://archive.is/TOuIQ#selection-4619.0-4619.241

Todd, B. (2020a). Misconceptions about effective altruism. *80,000 Hours*. https://80000hours.org/2020/08/misconceptions-effective-altruism/

Todd, B. (2020b). New data suggests the 'leaders" priorities represent the core of the community. *EA Forum*. https://forum.effectivealtruism.org/posts/DoAWvhgEnhSMr6yjS/new-data-suggests-the-leaders-priorities-represent-the-core

Todd, B. (2021). Is effective altruism growing? An update on the stock of funding vs people. *80,000 Hours* https://web.archive.org/web/20210803085830/https://80000hours.org/2021/07/effective-altruism-growing/

Todd, B. (2022). Comment on: Free-spending EA might be a big problem... *EA Forum*. https://forum.effectivealtruism.org/posts/HWaH8tNdsgEwNZu8B/free-spending-ea-might-be-a-big-problem-for-optics-and?commentId=3ThDJHqP9KqqjbkeE

Tomasik, B. (2006). Why Activists Should Consider Making Lots of Money [archived 5/17/2008]. *Utilitarian Essays*. https://web.archive.org/web/20080517114312/http://www.utilitarian-essays.com/make-money.html

Training for Good. (2022). *Tarbell Fellowship [archived 9/6/2022]*. https://web.archive.org/web/20220906201726/https://www.tarbellfellowship.org/fellowship

Transhumanismo.org. (2003). *Primera Reunión - Fundación de la Asociación Transhumanista Argentina* https://archive.is/FT5rg

Transhumanismo.org. (2004). *Segunda Reunión - Asociación Transhumanista Argentina*. https://web.archive.org/web/20200719183933/https://transhumanismo.org/old/ar/reuniones/reunion002.htm

User: AISafetyIsNotLongtermist. (2022). My Most Likely Reason to Die Young is AI X-risk. *EA Forum*. https://web.archive.org/web/20220706195154/https://forum.effectivealtruism.org/posts/hJDid3goqqRAE6hFN/my-most-likely-reason-to-die-young-is-ai-x-risk

User: akrolsmir. (2020). Comment on: Evidence, cluelessness, and the long term. *EA Forum*. https://forum.effectivealtruism.org/posts/LdZcit8zX89rofZf3/evidence-cluelessness-and-the-long-term-hilary-greaves?commentId=DEQJMnhBEETkJWrLS

User: Anonymous56. (2022, June 30). How to Prevent EA From Ever Turning Into a Cult. *EA Forum*. https://web.archive.org/web/20220810134757/https://forum.effectivealtruism.org/posts/5mNBM5xBTvugjqkag/how-to-prevent-ea-from-ever-turning-into-a-cult

User: Arepo [Sasha Cooper]. (2011, June 4). $100 for a new name for High Impact Careers. *Felicifia*. https://felicifia.github.io/thread/512.html

User: ArthurRainbow. (2020). AIRCS Workshop: How I failed to be recruited at MIRI. *LessWrong*. https://www.lesswrong.com/posts/EYhWxoaCK2YSpTei9/aircs-workshop-how-i-failed-to-be-recruited-at-miri

User: electroswing. (2022). Think about EA alignment like skill mastery, not cult indoctrination. *EA Forum*. https://web.archive.org/web/20220108011610/https://forum.effectivealtruism.org/posts/zZqFYXGhEKNuBD6va/think-about-ea-alignment-like-skill-mastery-not-cult

User: Gaverick [Jason Gaverick Matheny]. (2006, May 21). Comment on: Surplus money. *Felicifia*. https://web.archive.org/web/20200518115601/http://felicifia.blogspot.com/2006/10/surplus-money.html

User: Jeremy. (2022, September 1). Reply to comment on: Effective altruism is no longer the right name for the movement. *EA Forum*. https://web.archive.org/web/20220928154921/https://forum.effectivealtruism.org/posts/2FB8tK9da89qksZ9E/effective-altruism-is-no-longer-the-right-name-for-the-1?commentId=aCi92Epi9Ph9zuQS7

User: LadyMorgana. (2011). Re: lucky me (comment July 6, 2011 at 10:49am). *Felicifia*. https://web.archive.org/web/20110913135302/http://felicifia.org/viewtopic.php?p=2853#p3181:~:text=Just%20to%20let,offshoot%20from%20GWWC.

User: Michael_PJ. (2022). Comment on: EA Funds donation platform is moving to Giving What We Can. *EA Forum*. https://web.archive.org/web/20220530090733/https://forum.effectivealtruism.org/posts/3G9w9HfDusFw9LqxQ/ea-funds-donation-platform-is-moving-to-giving-what-we-can?commentId=B29YmmF6jFFKcKbai

User: nadavb. (2021). The case against "EA cause areas". *EA Forum*. https://forum.effectivealtruism.org/posts/3voXaqvPutSrHvuCT/the-case-against-ea-cause-areas

User: nadavb. (2022, October 1). EA may look like a cult (and it's not just optics). *EA Forum*. https://web.archive.org/web/20221126125529/https://forum.effectivealtruism.org/posts/7zg6pkrL6nQDLHPmk/ea-may-look-like-a-cult-and-it-s-not-just-optics

User: ParthThaya. (2020, September 8). How we promoted EA at a large tech company. *EA Forum*. https://forum.effectivealtruism.org/posts/BG68BvbaaT32u76kr/how-we-promoted-ea-at-a-large-tech-company-1

User: ParthThaya. (2022). Effective altruism is no longer the right name for the movement. *EA Forum*. https://web.archive.org/web/20220911050439/https://forum.effectivealtruism.org/posts/2FB8tK9da89qksZ9E/effective-altruism-is-no-longer-the-right-name-for-the-1

User: Raemon [Raymond Arnold]. (2012a, August 8). The High Impact Network (THINK) - Launching Now. *LessWrong*. https://web.archive.org/web/20131224214016/http://lesswrong.com/lw/dyj/the_high_impact_network_think_launching_now

User: Raemon [Raymond Arnold]. (2012b, August 8). Reply to comment on: The High Impact Network (THINK) - Launching Now. *LessWrong*. https://web.archive.org/web/20131231085649/http://lesswrong.com/lw/dyj/the_high_impact_network_think_launching_now/76bj

User: Raemon [Raymond Arnold]. (2019). Comment on: Is EA Growing? EA Growth Metrics for 2018. *EA Forum*. https://web.archive.org/web/20190703044222mp_/https://forum.effectivealtruism.org/posts/MBJvDDw2sFGkFCA29/is-ea-growing-ea-growth-metrics-for-2018#agBw2qp57oBLPfuRF

User: Raemon [Raymond Arnold]. (2022, December 9). Reply to comment on: Why did CEA buy Wytham Abbey? *EA Forum*. https://archive.vn/90tRF

User: SeeYouAnon. (2022). Reply to Pietsch on: Why did CEA buy Wytham Abbey? *EA Forum*. https://web.archive.org/web/20221210195553/https://forum.effectivealtruism.org/posts/xof7iFB3uh8Kc53bG/why-did-cea-buy-wytham-abbey?commentId=AS3fnqR9weqtL7iDC#:~:text=Just%20for%20clarity%27s,deal%20it%20is.

User: utilitymonster. (2011a, June 22). Comment at 2:11pm on: Making a Utilitarian Charity. *Felicifia*. https://felicifia.github.io/thread/497.html#p4242

User: utilitymonster. (2011b, June 22). Comment at 8:51pm on: Making a utilitarian charity. *Felicifia*. https://felicifia.github.io/thread/497.html#p4249

User: utilitymonster. (2012, June 4). Comment on: Why people give to developing-world charities. *Felicifia*. https://felicifia.github.io/thread/629.html#p5640

User: weeatquince [Sam Hilton]. (2020). Comment on: What will 80,000 provides and not provide within the EA community. *EA Forum*. https://archive.is/yFyDr

Vaughan, K. (2017, May 12). The fidelity model of spreading ideas. *Centre for Effective Altruism*. https://www.centreforeffectivealtruism.org/blog/the-fidelity-model-of-spreading-ideas/

Vollmer, J. (2018). Comment on: Max Dalton: I'm pleased to announce the second edition of the effective altruism handbook. *Effective Altruism - Facebook Group (Public)*. https://web.archive.org/web/20200805114604/https://www.facebook.com/groups/effective.altruists/permalink/1750780338311649/?comment_id=1751075288282154&reply_comment_id=1753905907999092

Vollmer, J. (2021). Some quick notes on "effective altruism". *EA Forum*. https://forum.effectivealtruism.org/posts/h566GT4ECfJAB38af/some-quick-notes-on-effective-altruism

Whittlestone, J. (2015). Career reviews: Research into risks from artifiical intelligence [archived 8/29/2018]. *80,000 Hours*. https://web.archive.org/web/20180829142906/https://80000hours.org/career-reviews/artificial-intelligence-risk-research/

Wiblin, R. (2009a, October 12). Do Singularitarians worry about their retirement? [archived 5/20/2020]. *Rob Wiblin*. https://archive.md/MVRHE

Wiblin, R. (2009b, October 12). Non-existent people are people too. *Rob Wiblin*. https://archive.md/LKITk

Wiblin, R. (2010a, May 27). Existential Risk Links. *Rob Wiblin's Old Blog*. https://oldblog.robwiblin.com/2010/04/18/existential-risk-links/

Wiblin, R. (2010b, May 27). Geoengineering as speculative existential risk. https://oldblog.robwiblin.com/2010/03/29/geoengineering-as-speculative-existential-risk/

Wiblin, R. (2010c). Why existential risk is the most important thing (Part I). *Rob Wiblin's Old Blog*. https://archive.st/archive/2020/5/oldblog.robwiblin.com/7s05/oldblog.robwiblin.com/2010/03/26/why-existential-risk-is-the-most-important-thing-part-i/index.html

Wiblin, R. (2010d). Why improve nature when destroying it is so much easier? [archived 6/30/2014]. *Rob Wiblin*. https://archive.md/HbE2a

Wiblin, R. (2012, May 27). Nick Bostrom interview on x-risk in The Atlantic. *Rob Wiblin's Old Blog*. http://archive.is/FaghI

Wiblin, R. (2015a, March 10). Disagreeing about what's effective isn't disagreeing with effective altruism. *80,000 Hours*. https://web.archive.org/web/20150918161936/https://80000hours.org/2015/07/disagreeing-about-whats-effective-isnt-disagreeing-with-effective-altruism/

Wiblin, R. (2015b). Where should you donate to have the most impact during giving season 2015? [archived 6/27/2016]. *80,000 Hours blog*. https://web.archive.org/web/20160627024525/https://80000hours.org/2015/12/where-should-you-donate-to-have-the-most-impact-in-giving-season-2015/

Wiblin, R., Harris, K., & MacAskill, W. (2020). Will MacAskill on the moral case against ever leaving the house, whether now is the hinge of history, and the culture of effective altruism. *80,000 Hours Podcast*. https://80000hours.org/podcast/episodes/will-macaskill-paralysis-and-hinge-of-history/

Wise, J., Mayhew, S., & Gertler, A. (2020). Advice for Responding to Journalists. https://docs.google.com/document/d/1jU4snbnAIq-q4Dl_mIF0JyTT_bQrvS1wdBhQC1H-Bv8/edit

Wise, J., & Sasi, D. (2022). *Context on effective altruism for mental health providers*. https://archive.is/as5xd

Woodside, T. (2022a). *About*. https://web.archive.org/web/20221130115846/https://www.thomaswoodside.com/about/

Woodside, T. (2022b). Comment on: EA and the current funding situation. *EA Forum*. https://forum.effectivealtruism.org/posts/cfdnJ3sDbCSkShiSZ/ea-and-the-current-funding-situation?commentId=oQPc8Wx5eMyrpEey6

Woodside, T., Hendrycks, D., Mazeika, M., Zhang, O., Hough, S., & Liu, K. (2022). Introducing the ML Safety Scholars Program. *EA Forum*. https://forum.effectivealtruism.org/posts/9RYvJu2iNJMXgWCBn/introducing-the-ml-safety-scholars-program

Yudkowsky, E. (2007). Expecting Short Inferential Distances. *LessWrong*. https://www.lesswrong.com/posts/HLqWn5LASfhhArZ7w/expecting-short-inferential-distances

Yudkowsky, E. (2013). Comment on: A Critique of Effective Altruism. *LessWrong*. https://www.lesswrong.com/posts/E3beR7bQ723kkNHpA/a-critique-of-effective-altruism#37WQGqigivLAxGrFR

Yudkowsky, E. (2015, August 11). *Reply to Rob Wiblin "Existential-risk focussed people should be more polite..."* https://web.archive.org/web/20190429003757/https://www.facebook.com/robert.wiblin/posts/693677434745?comment_id=693682095405&offset=0&total_comments=21&comment_tracking={%22tn%22%3A%22R6%22}

Secondary Sources

Aleem, Z. (2022, December 3). How Sam Bankman-Fried's fall exposes the perils of effective altruism. *MSNBC*. https://www.msnbc.com/opinion/msnbc-opinion/ftx-sbf-effective-altruism-bankman-fried-rcna59172

Bajekal, N. (2022, August 10). Want to Do More Good? This Movement Might Have the Answer. *Time*. https://time.com/magazine/us/6206429/august-22nd-2022-vol-200-no-7-u-s/

Berg, A. (2018). Effective Altruism: How Big Should the Tent Be? *Public Affairs Quarterly*, 32(4), 269-287.

Berkey, B. (2018). The Institutional Critique of Effective Altruism. *Utilitas*, 30(2), 143-171.

Bowen, G. A. (2008). Naturalistic inquiry and the saturation concept: a research note. *Qualitative Research*, 8(1), 137-152.

Bowen, G. A. (2009). Document Analysis as a Qualitative Research Method. *Qualitative Research Journal*, 9(2), 27-40. https://doi.org/10.3316/QRJ0902027

Brest, P. (2015, July 1). The Logic of Effective Altruism - Forum Response. *The Boston Review*. http://bostonreview.net/forum/logic-effective-altruism/paul-brest-response-effective-altruism

Brown, S. G. (2016). Supporting the best charities is harder than it seems. *Journal of Global Ethics*, 12(2), 240-244. https://doi.org/10.1080/17449626.2016.1205118

Burton, P. (2015, October 13). *Family Gives Away Half Their Income to Help Others*. https://boston.cbslocal.com/2015/10/13/family-gives-away-half-their-income-to-help-others/

Chapman, P. (2011, February 13). 3 Grad Students at Rutgers Pledge Lifetime of Paychecks to Lifesaving Charities. *The Chronicle of Higher Education*. https://www.chronicle.com/article/3-Grad-Students-at-Rutgers/126019

Clough, E. (2015, July 14). Effective Altruism's Political Blind Spot. *Boston Review*. http://bostonreview.net/world/emily-clough-effective-altruism-ngos

Cochrane, L., & Thornton, A. (2016). Charity rankings: Delivering development or dehumanising aid? *Journal of International Development*, 28(1), 57-73.

Côté, N., & Steuwer, B. (2023). Better vaguely right than precisely wrong in effective altruism: the problem of marginalism. *Economics & Philosophy*, 39(1), 152-169.

Crary, A. (2021). Against 'Effective Altruism'. *Radical Philosophy*, 2.10(Summer 2021), 33-43. https://www.radicalphilosophy.com/wp-content/uploads/2021/07/rp210_crary.pdf

Deaton, A. (2015, July 1). The Logic of Effective Altruism: Forum Response. *Boston Review*, *July/August 2015*. http://bostonreview.net/forum/logic-effective-altruism/angus-deaton-response-effective-altruism

Douglas, D. M. (2016). Doxing: a conceptual analysis. *Ethics and Information Technology*, 18, 199-210. https://doi.org/10.1007/s10676-016-9406-0

Eikenberry, A. M., & Mirabella, R. M. (2018). Extreme Philanthropy: Philanthrocapitalism, Effective Altruism, and the Discourse of Neoliberalism. *PS: Political Science & Politics*, 51(1), 43-47. https://doi.org/10.1017/S1049096517001378

Fisher, A. (2022, September 22). Sam Bankman-Fried Has a Savior Complex - And Maybe You Should Too. *Sequoia Capital*. https://web.archive.org/web/20220922165220/https://www.sequoiacap.com/article/sam-bankman-fried-spotlight/

Foucault, M. (1978). *The History of Sexuality, Volume 1: An Introduction* (R. Hurley, Trans.). Pantheon Books.

Gabriel, I. (2017). Effective Altruism and its Critics. *Journal of Applied Philosophy*, 34(4), 457-473.

Heyboer, K. (2010, December 3). Rutgers University students launch movement to inspire lifelong

charity. *NJ.com*. https://www.nj.com/news/2010/12/rutgers_university_students_or.html

Hine, C. (2017). Ethnography and the Internet: Taking Account of Emerging Technological Landscapes. *Fudan Journal of the Humanities and Social Sciences*, 10(3), 315-329. https://doi.org/10.1007/s40647-017-0178-7

Hookway, N., & Snee, H. (2017). The Blogosphere. In N. G. Fielding, R. M. Lee, & G. Blank (Eds.), *The Sage Handbook of Online Research Methods* (pp. 380-398). Sage.

Hossenfelder, S. (2022). Elon Musk & The Longtermists: What Is Their Plan? *Sabine Hossenfelder - Science Without the Gobbledygook*. https://youtu.be/B_M64BSzcRY

Humphrey, S. (2004). No death, please, I'm bionic. *NOW Toronto*. https://nowtoronto.com/news/no-death-please-im-bionic

Igneski, V. (2016). Living a meaningful and ethical life in the face of great need: Responding to Singer's The Most Good You Can Do. *Journal of Global Ethics*, 12(2), 147-153. https://doi.org/10.1080/17449626.2016.1191036

Impett, L. (2018). Prometheus Wired. *New Left Review*, *May/June* 2018, 151-159.

Isaacs, T. (2016). The most good *we* can do: comments on Peter Singer's *The Most Good You Can Do*. *Journal of Global Ethics*, 12(2), 154-160. https://doi.org/10.1080/17449626.2016.1191037

Janetzko, D. (2017). Nonreactive Data Collection Online. In N. G. Fielding, R. M. Lee, & G. Blank (Eds.), *The Sage Handbook of Online Research Methods* (pp. 76-91). Sage.

Joachim, J., & Schneiker, A. (2012). New humanitarians? Frame appropriation through private military and security companies. *Millennium*, 40(2), 365-388.

Keim-Malpass, J., Steeves, R. H., & Kennedy, C. (2014). Internet ethnography: A review of methodological considerations for studying online illness blogs. *International journal of nursing studies*, 51(12), 1686-1692.

Kissel, J. (2017). Effective altruism and anti-capitalism: an attempt at reconciliation. *Essays in Philosophy*, 18(1), 5.

Krishna, N. (2016). Add Your Own Egg. *The Point*. https://thepointmag.com/2016/examined-life/add-your-own-egg

Kurzy.cz. (2022). *Hostačovský zámek s.r.o. , Skryje IČO 27798836 - Obchodní rejstřík firem*. https://archive.vn/xLeSP

Law, K. F., Campbell, D., & Gaesser, B. (2021). Biased Benevolence: The Perceived Morality of Effective Altruism Across Social Distance. *Personality and Social Psychology Bulletin*, 48(3), 1-19. https://doi.org/10.1177/01461672211002773

Lechterman, T. M. (2020). The Effective Altruist's Political Problem. *Polity*, 52(1), 88-115. https://www.academia.edu/37041651/The_Effective_Altruists_Political_Problem

Lee, A. M., & Lee, E. B. (1939). *The Fine Art of Propaganda: A Study of Father Coughlin's Speeches*.

Lenman, J. (2022). How effective altruism lost the plot: Why longtermism is not the solution. *iai News*. https://iai.tv/articles/how-effective-altruism-lost-the-plot-auid-2284

Levitz, E. (2022, August 30). Why Effective Altruists Fear the AI Apocalypse: A conversation with the philosopher William MacAskill. *New York Magazine - Intelligencer*. https://nymag.com/intelligencer/2022/08/why-effective-altruists-fear-the-ai-apocalypse.html

Lewis-Kraus, G. (2022). The Reluctant Prophet of Effective Altruism. *The New Yorker*. https://www.newyorker.com/magazine/2022/08/15/the-reluctant-prophet-of-effective-altruism

Lim, C.-M. (2019). Effectiveness and Ecumenicity. *Journal of Moral Philosophy*, 16(5), 590-612.

MacFarquhar, L. (2015). *Strangers Drowning: Voyages to the Brink of Moral Extremity*. Penguin UK.

Matthews, D. (2013, May 31). Join Wall Street. Save the World. *Washington Post*. https://www.washingtonpost.com/news/wonk/wp/2013/05/31/join-wall-street-save-the-world/

Matthews, D. (2022). How effective altruism went from a nice movement to a billion-dollar force. *Vox - Future Perfect*. https://www.vox.com/future-perfect/2022/8/8/23150496/effective-altruism-sam-bankman-fried-dustin-moskovitz-billionaire-philanthropy-crytocurrency

Melucci, A. (1995). The Process of Collective Identity. In H. Johnston & B. Klandermans (Eds.),

Social Movements and Culture (pp. 41-63). University of Minnesota Press. http://www.jstor.org/stable/10.5749/j.ctttt0p8.6

Mihailov, E. (2022). Measuring Impartial Beneficence: A Kantian Perspective on the Oxford Utilitarianism Scale. *Review of Philosophy and Psychology*, 1-17. https://doi.org/10.1007/s13164-021-00600-2

MSNBC. (2010, December 21). *The Dylan Ratigan Show - interview with Nick Beckstead.*

Muyskens, K. (2017). The Other Half of Effective Altruism: Selective Asceticism. *Essays in Philosophy*, *18*(1), 6.

Naughton, J. (2022, December 4). Longtermism: how good intentions and the rich created a dangerous creed. *The guardian*. https://www.theguardian.com/technology/commentisfree/2022/dec/04/longtermism-rich-effective-altruism-tech-dangerous

Ongweso Jr., E. (2022, November 23). Ok, WTF is 'Longtemism', the Tech Elite Ideology That Led to the FTX Collapse? *Vice*. https://www.vice.com/en/article/bvmanv/ok-wtf-is-longtermism-the-tech-elite-ideology-that-led-to-the-ftx-collapse

Pellegrino, G. (2017). Effective Altruism and the Altruistic Repugnant Conclusion. *Essays in Philosophy*, *18*(1), 4.

Pinker, S. [@sapinker]. (2022). *Why 'longtermism' isn't ethically sound by @ChristineEmba - I'm not the only one concerned that the Effective Altruism movement may lurch toward hypothetical helping of the nonexistent, neither E nor A.* Twitter. https://twitter.com/sapinker/status/1567143610952388608

Pinto, A. T. (2019). Capitalism with a Transhuman Face. *Third Text*, 1-22. https://doi.org/10.1080/09528822.2019.1625638

Read, R. (2018). Must Do Better. *Radical Philosophy*, *201*(February 2018), 106-109. https://www.radicalphilosophy.com/reviews/individual-reviews/must-do-better

Roache, C. (2010, December 8). Group cites effects of charity donations. *The Daily Targum (Rutgers University)*, 5. https://issuu.com/targum_editor/docs/dt_2010-12-08

Rubenstein, J. C. (2016). The Lessons of Effective Altruism: Review of *The Most Good You Can Do* (Peter Singer) and *Doing Good Better* (William MacAskill). *Ethics & International Affairs*, 30(4), 511-526. https://doi.org/doi:10.1017/S0892694160004 84

Schaller, W. E. (2018). Must we maximize the good when we do good? *Journal of the Texas Tech Ethics Center*, 2(1), 44-59. https://ttuec-ojs-ttu.tdl.org/ttuec/index.php/ttuec/article/view/15

Schuster, J., & Woods, D. (2021). *Calamity Theory: Three Critiques of Existential Risk*. University of Minnesota Press.

Skelton, A. (2016). The ethical principles of effective altruism. *Journal of Global Ethics*, 12(2), 137-146. https://doi.org/10.1080/17449626.2016.1193552

Smithey, L. A. (2009). Social movement strategy, tactics, and collective identity. *Sociology Compass*, 3(4), 658-671.

Snow, D. A. (2001). *Collective Identity and Expressive Forms* (CSD Working Paper, Issue. UC Irvine: Center for the Study of Democracy. https://escholarship.org/uc/item/2zn1t7bj

Snow, D. A. (2004). Framing Processes, Ideology, and Discursive Fields. In D. A. Snow, S. A. Soule, & H. Kriesi (Eds.), *The Blackwell Companion to Social Movements*. Blackwell Publishing, Ltd.

Srinivasan, A. (2015, Sept 24). Stop the Robot Apocalypse. *London Review of Books*, 37(18), 3-6. https://www.lrb.co.uk/v37/n18/amia-srinivasan/stop-the-robot-apocalypse

Štorkán, M. (2022). Česká stopa zkrachovalé burzy FTX: nadace krále kryptoměn poslala přes sto milionů na nákup zámku [article text translated using Google Translate]. *iRozhlas*. https://www.irozhlas.cz/zpravy-domov/ftx-krypto-burza-bankman-fried-kryptomeny_2212060500_sto

Syme, T. (2019). Charity vs. Revolution: Effective Altruism and the Systemic Change Objection. *Ethical Theory and Moral Practice*, 1-28.

The Economist. (2018, May 31). To help save the world, become a banker not a doctor. *The Economist*. https://www.economist.com/international/2018/05/31/to-help-save-the-world-become-a-banker-not-a-doctor

Torres, P. (2020). *Transhumanism, Effective Altruism, and System Theory: A Brief History and Analysis*

of Existential Risk Research. Global Challenges. https://globalchallenges.org/wp-content/uploads/Transhumanism-Effective-Altruism-and-Systems-Theory-2-1.pdf

Unger, P. (1996). *Living High and Letting Die: Our Illusion of Innocence*. Oxford University Press.

Van Dijk, T. A. (2006). Ideology and Discourse Analysis. *Journal of Political Ideologies*, 11(2), 115-140. https://doi.org/10.1080/13569310600687908

Visram, T. (2022, September 14). The truth about Elon Musk, Sam Bankman-Fried, and effective altruism. *Fast Company*. https://www.fastcompany.com/90784622/the-truth-about-elon-musk-sam-bankman-fried-and-effective-altruism

Whyte, J. (2014). Humanizing Militarism: Amnesty International and the Tactical Polyvalence of Human Rights Discourses. In A. Yeatman & P. Birminghan (Eds.), *The Aporia of Rights* (pp. 183-204). Bloomsbury.

Whyte, J. (2019). *The Morals of the Market: Human Rights and the Rise of Neoliberalism*. Verso.

Wimmer, R. D., & Dominick, J. R. (2014). *Mass Media Research: An Introduction* (Tenth ed.). Wadsworth.

Wolf, L. (2022). Regency Centers transforms iconic Harvard Square buildings into The Abbot, a vibrant mixed-use project. *REIT Magazine*, *September/October* 2022. https://www.reit.com/news/reit-magazine/september-october-2022/regency-centers-transforms-iconic-harvard-square

Yanow, D. (2006). Neither Rigorous nor Objective? Interrogating Criteria for Knowledge Claims in Interpretive Science. In D. Yanow & P. Schwartz-Shea (Eds.), *Interpetation and Method: Empirical Research Methods and the Interpretive Turn* (pp. 67-88). M.E. Sharpe.

Yetter Chappell, R. (2010). GWWC Rutgers Launch. *Philosophy, et cetera*. https://www.philosophyetc.net/2010/12/gwwc-rutgers-launch.html#more

Zhang, L. (2016, July 14). This Startup Entrepreneur Lives on Min. Wage So He Has More Money to Donate. *Huffington Post*. https://www.huffpost.com/entry/this-startup-entrepreneur-lives-on-minimum-wage-so_b_5783da0ee4b05b4c02fd2e0b

Ziatchik, A. (2022, October 24). The Heavy Price of Longtermism. *The New Republic*. https://newrepublic.com/article/168047/longtermism-future-humanity-william-macaskill

University of Antwerp
IOB | Institute of
Development Policy