

The limits of quantified compassion: A critique of effective altruism

Martin Foltin¹

Abstract

This paper critically evaluates effective altruism (EA) and contends that its utilitarian methodology undercuts its compassionate aims. The analysis identifies three major critiques: First, EA's treatment of humans as purely rational agents overlooks motivational complexity, unstable preferences, and the intrinsic moral value of relationships, labeling these factors as cognitive biases. Second, EA's strong focus on measurability reduces well-being to quantifiable metrics such as QALYs, neglecting other significant values, such as dignity, community, and the potential for systemic progress. This approach presents ideological decisions as mere objective calculations. Third, EA reflects neoliberal tendencies by depoliticizing structural issues and presenting justice as a matter of individual philanthropy, notably illustrated in *Earning to Give*. This study maintains that EA conflates altruism with 'effective philanthropy'—merely a rationalized system of distributing resources. We propose adopting methodological pluralism to reincorporate emotion, social context, and political participation into moral reasoning, with an emphasis on personal engagement in systemic change over calculation-driven, individualistic methods.

Keywords: effective altruism, utilitarianism, philanthropy, neoliberalism, measurability, moral decision-making, systemic change

Introduction

Imagine a young graduate faced with a life decision. She receives a lucrative job offer from an investment bank in a distant financial center. This job would let her donate tens of thousands of euros a year to save hundreds of anonymous children in Africa from malaria. On the other hand, she can stay at home to care for her ailing mother, offering emotional support and maintaining community ties. These bonds cannot be quantified in a spreadsheet. Effective altruism (EA) gives an unyielding answer. Building on Singer's essay, *Famine, affluence, and morality* (Singer, 1972), EA presents a vision of morality free from sentiment and geography. Its promise is simple: use reason, evidence, and calculation to do the maximum good with our resources (MacAskill, 2015b).

Singer's principle (1972, p. 231) states: if we can prevent something bad without sacrificing anything of similar moral importance, we ought to do so. This remains central to the movement. Singer (2015, p. 103) later reframes altruism as a win-win: donors gain meaning, recipients gain life. This is mainly a motivational strategy, not a true revision of values. By making altruism emotionally rewarding, EA risks weakening the moral weight of altruistic acts. The donor chooses an "investment" but remains distant from the moral impact on those not helped. Here, value is measured not by 'warm glow' (Elster, 2011), but by quantifiable metrics like lives saved or QALYs—quality-adjusted life years that combine life expectancy with quality of life.

EA promises to fulfill empathy through reason. This isn't a claim that effective altruists lack compassion. Rather, their utilitarian methodology is so reductionist that it distorts their original empathy. The attempt to convert solidarity into science yields a paradox: maximizing good can cause us to lose sight of what is truly good.

The goal of this article is to examine closely the reductionist human model underpinning EA. We claim that the scientific approach to altruism produces three core problems: it misconstrues the nature of human rationality and motivation, dismissing the depths of preferences and relationships; it creates a dependence on measurable metrics, leading to the neglect of unquantifiable but crucial values such as dignity, community, and systemic change; and it adopts a detached and abstract view of morality, missing the importance of social and

¹ Institute of Philosophy of the Slovak Academy of Sciences, v.v.i., Bratislava (Slovakia); email: filomafo@savba.sk; ORCID: 0000-0003-3981-7811

relational contexts. These three themes form the structure of our critique.

The following analysis demonstrates that, beneath the appearance of scientific objectivity and rational calculation, EA exhibits epistemic overconfidence and advances a particular ideology. This framework may ultimately weaken authentic human solidarity. Rather than offering moral certainty, EA oversimplifies complex moral questions and, by trying to maximize efficiency, risks losing touch with the core meaning of altruism.

This analysis uses three main source types: foundational texts (Singer, 1972, 2011, 2015; MacAskill, 2015a, 2015b), official documents, and data from GiveWell (2015, 2019, 2025). It includes IDinsight's 2019 study on beneficiary preferences and representative texts from the *EA Handbook* (Alexander, 2015; Yudkowsky, 2015). The selection analyzes EA within its own framework and focuses on central texts and mainstream practices.

The myth of the rational actor and the problem of preferences

EA advocates may say their model is normative, not descriptive – it prescribes how people should act morally. Yet demanding pure rationality faces three challenges, each of which undermines utilitarian maximization.

The first challenge comes from the moral value of relationships and place-based commitments. Economists Banerjee and Duflo (2019) note that the logical idea of mass migration for economic gains seems “simple, seductive, and wrong”. According to economic rationality, a steelworker from Pennsylvania should move to find work. Yet, empirical studies show the opposite: people stay put, facing job loss and social decline. This leads to phenomena like *deaths of despair* (Banerjee & Duflo, 2019; De Wispelaere, 2004, p. 16). Why? Decisions are influenced by factors that are hard to quantify, such as family ties, community, comfort, fear of the unknown, and dignity.

If people weren't rational egoists for themselves, why would they be in altruistic acts? Green (2008, p. 110) observes that attitudes, beliefs, and relationships shape behavior as much as self-interest. This relational embeddedness isn't biased but core to moral life. Utilitarian maximization misses this, seeing obligations rooted more in abstract principle than real relationships (Liu, 2025).

Singer does not soften his stance in later years. In his 2017 defense of utilitarianism, he endorses William Godwin's ‘fire’ thought experiment, arguing that “justice would have taught me to save the life of [a benefactor to humanity] at the expense of the other [even if she were my mother]” (de Lazari-Radek & Singer, 2017, p. 79). For Singer, the pronoun *may* not overturn the “decisions of everlasting truth” (de Lazari-Radek & Singer, 2017, p. 79). However, this position faces resistance. Crisp (2025, p. 205) argues that most rational beings would reject such impartiality, challenging utilitarianism's claim to intuitive appeal. This suggests that EA's methodology goes beyond efficiency; it deliberately overrides the relational basis of morality in favor of a universal utility viewpoint (Singer & de Lazari-Radek, 2014). This view treats people as utility locations, not subjects within a broader value structure that includes humanity and dignity (Gluchman, 2017).

This demand for rational maximization is problematic not only because it is empirically unfeasible. More importantly, it fails due to ontological reductionism. The shift from descriptive psychology to normative critique is not a categorical error. Instead, it shows that EA functions as a form of governmentality, reshaping and ultimately dissolving the moral subject. By reducing individuals to comparable utility units, EA turns people into objects, stripping them of their inherent dignity and individual recognition. If theory replaces ethical responsibility with market logic, it becomes a tool for management rather than ethics. A theory that destroys ethical perception of “good” by replacing it with calculation is internally contradictory and philosophically untenable.

The next challenge: preferences are unstable and vague. Banerjee and Duflo (2019, p. 165)

ask, if we don't know our feelings on trivial choices, how can we have clear preferences about climate change? Skepticism about the precise specification of preferences, as seen in Hammond (1982, p. 101) and Mirrlees (1982, p. 67), highlights the shaky foundation of utilitarian calculation. If stable preferences don't exist, what is being maximized?

Utilitarianism sees people as “locations of utility”, merely sites where pleasure or pain occurs (Sen & Williams, 1982, pp. 4–5). After noting someone's utility, utilitarianism wants no more information. This view reduces people to passive utility carriers, ignoring motivational complexity—something economists rarely seek insight into from other disciplines (Green, 2008, p. 434).

The third challenge concerns adaptive preferences, which Elster analyzes. His key question is: “Why should the satisfaction of individual desires be the criterion of justice and social choice when individual desires themselves may be shaped by a process that precedes this choice?” (Elster, 1982, p. 219; Schweiger & Graf, 2015, p. 17). Imagine a community living under long-term oppression, whose members, as a result of the *sour grapes* mechanism, have adapted their desires so that they only want what is achievable within the given system. On a hedonistic scale, their level of satisfaction may be indistinguishable from that of a person with fully autonomous desires. For a strict utilitarian who measures only the resulting utility, there may be no moral problem here. However, this approach fails to distinguish between happiness resulting from resignation and that resulting from autonomous fulfillment, which, paradoxically, can justify the *status quo* of oppression by treating constrained choices as genuine preferences worthy of maximization.

The illusion of measurability and the tyranny of the quantifiable

The problem with EA does not end with the (ir)rationality of the subject but continues at the level of the decision-making process itself. EA is based on the belief that we are not only capable of acting rationally, but also of accurately measuring and comparing the consequences of our actions in order to identify the most effective option. However, this assumption runs up against a fundamental distinction between quantifiable risk and radical uncertainty (so-called *Knightian uncertainty*), in which we do not even know the probabilities of outcomes (Banerjee & Duflo, 2019). We are condemned to freedom and responsibility, but our decisions are inevitably made under conditions of limited knowledge.

Consider the dilemma of the effective altruist: should they move away from their aging mother to maximize their income and donations to save lives in Africa, or should they stay and care for her? While EA offers a clear calculation for the first option and may condemn staying as a moral failure based on that calculation, the second option is not just a sentimental choice. It is an investment in immeasurable but crucial social capital. Strengthening community ties, dignity, family cohesion, education, and cultural identity creates an environment from which innovation, resilience, and long-term prosperity can emerge. However, these dimensions of human experience are excluded from the analysis because they defy simple quantification. The pursuit of efficiency thus leads to the *tyranny of the measurable* and the systematic neglect of complex solutions—from supporting investigative journalism to strengthening local democracy or advocating systemic debt relief and financial regulation (Kalajtidis, 2025).

The practice of GiveWell, the movement's flagship organization, perfectly illustrates this effort to turn ethics into a technical problem. Although EA advocates argue that an approach based on transparent “moral weights”—numerical values assigned to compare the relative importance of helping different groups or causes—is a sign of intellectual honesty—better an imperfect but explicit model than intuition—in reality, it is epistemic arrogance. It presents deeply subjective and value-laden estimates as objective, almost scientific truth, analogous to the *black-box* problem in artificial intelligence ethics (Karastergiou, 2025, p. 114). As Batôt and Belli (2025, p. 199) argue in the context of medical AI, such algorithmic proxies often bypass

the subjective and deliberative processes that confer moral authority on preferences, treating individuals as mere collections of signals. This phenomenon is not limited to philanthropy. Shaw and Vanadia (2022) have named it the *hidden assumptions problem* in the context of crisis public policy (pp. 73–75). GiveWell’s methodology leads to a similar simplification, in which good is commodified and reduced to a binary state: alive or dead.

This reductionism is not a misinterpretation by critics; it is explicitly advocated as “conscious reasoning” over “automatic [emotional] response” (de Lazari-Radek & Singer, 2017, p. 36). Singer maintains that, in public policy, well-being must be quantified in units such as Quality-Adjusted Life-Years (QALYs), which “can be added, subtracted, multiplied, and divided” (de Lazari-Radek & Singer, 2017, pp. 72–73). However, this mathematical certainty is an epistemological mirage. As Ng (2020) points out, EA faces the *second-best challenge*: in a complex world where indirect effects are vast and unmeasurable, focusing solely on a single “measurable” intervention (such as malaria nets) without accounting for systemic repercussions can lead to a “nihilistic conclusion”. If we cannot calculate the long-term ripple effects, the “rational maximization” of EA ceases to be a normative ideal and becomes a mere “third-best policy”—a guess based on information scarcity (Ng, 2020). When the value of a charity is “imprecise”, the entire imperative to find the “single best” organization collapses, turning EA from an exact science into a speculative, and often biased, philanthropy. As Baranová (2025) points out in the context of neonatal care, such utilitarian frameworks often falter when faced with genuine prognostic uncertainty, as individual human trajectories frequently deviate from the statistical expectations required for “rational” bioethical calculation.

The critique here is not that lack of measurement leads to moral invalidity *per se*, but rather that the “measurable bias” functions as a normative filter that distorts the moral landscape. By elevating epistemic certainty as a prerequisite for moral action, EA commits a “category error”: it treats the difficulty of quantification as a lack of moral value. As Kissel (2017) argues, EA’s focus on *marginal impact*—what an individual can change immediately—systematically ignores collective political action. This is not a mere technical oversight; it is a normative failure. This “technocratic dogmatism” prioritizes short-term palliative interventions over long-term structural cures because the latter cannot be captured by the methodology of randomized controlled trials (RCTs). Consequently, EA does not just measure good; it defines “good” in a way that excludes systemic justice from the outset.

As Henriquez (2016, p. 25) shows, based on GiveWell’s recommended allocation as of August 2015, up to 87% of the funds went to life-saving interventions, while only 13% went to life-improving interventions. This trend has not only persisted but intensified: current data from 2024 demonstrates that all four programs designated as *Top Charities* remain exclusively focused on saving children’s lives—Against Malaria Foundation (mosquito nets), Malaria Consortium (preventive malaria treatment), Helen Keller International (vitamin A), and New Incentives (vaccination incentives). Not a single top-recommended program is primarily aimed at improving quality of life, systemic change, or unmeasurable values such as dignity or community (GiveWell, 2025). Human life thus becomes an item on the global philanthropic market, where donors seek the best “return” on their investment, as suggested by the comparison that “your dollar has more value”—for example, \$3,000–5,500 to save a life in Africa versus significantly higher costs for less measurable assistance to a child in the US (GiveWell, 2025).

The process of “producing” these moral figures takes on absurd proportions when we examine a 2019 study commissioned by GiveWell to determine the preferences of aid recipients in Ghana and Kenya. Under the guise of an inclusive approach, what actually took place was an act of ethical outsourcing and methodological violence. Responsibility for brutal moral compromises was shifted to respondents living in extreme poverty, who were asked questions that transformed complex moral dilemmas into technical calculations, including questions

about willingness to pay for mortality risk reduction or choices between different combinations of lives saved and financial transfers (IDinsight, 2019, pp. 18–22). Moral decision-making was thus transformed into a laboratory experiment aimed at giving the company an aura of scientific objectivity through absurdly precise figures such as “the value of a statistical life is \$40,763” (IDinsight, 2019, p. 20; GiveWell, 2019).

However, the study revealed more than its authors intended. First, it exposed the glaring difference between the values of Western experts and the people they are supposed to help. While the median “moral weight” of GiveWell employees assigned a value to the life of a child under 5 that was almost half that of an older person (ratio of 0.55), respondents in Ghana and Kenya consistently expressed the exact opposite (ratio of 1.2 to 1.9). This proves that the quantitative approach is not a neutral tool, but a carrier of culturally conditioned prejudices (IDinsight, 2019). Secondly, the most revealing moment came when 38% of respondents refused to play by the rules, arguing that life cannot be compared to money. However, this fundamental philosophical opposition to the premise of quantification did not prompt questioning of the framework. On the contrary, it was coldly incorporated as another data point into the calculation. The protest against the system thus became the fuel that kept it running.

The attempt to fit human life into a regression equation led to absurd results, such as that the value of a human life over the age of 40 is negative – a result that the study authors themselves expressed low confidence in, stating they were “not confident that the negative estimates for those above 40 reflect true preferences based on the mechanics of the analytical model” (IDinsight, 2019, p. 32). However, the ultimate cynicism was revealed in the conclusion: after this ethically questionable research, GiveWell changed its moral weights, but with disarming honesty added that this change had almost no impact on the actual distribution of funds. The entire research thus proved to be not a tool for seeking truth, but an act of performative rationality aimed at strengthening the organization’s authority as the entity entitled to perform these complex calculations.

The crisis of methodology runs even deeper. The GiveWell study itself showed that different quantitative approaches lead to conflicting conclusions: the preference lens favors the life of a child, while the future economic value model arrives at the opposite conclusion (IDinsight, 2019). The claim that we are guided by “data” becomes an empty phrase, since the choice of measurement method itself is an ideological choice. In the context of developing nations, such a framework proves inadequate for designing equitable public health policies, as it privileges abstract interests over communal norms and traditional morality (Dimonye et al., 2025). Moreover, even if we agree on a method, we face the unsolvable dilemma of utilitarianism: should we maximize total welfare or average welfare? As Greaves shows, these principles lead to radically different conclusions, and the choice between them is not a technical one, but a profound value choice with no objective solution (Greaves, 2019, pp. 52–53).

The pursuit of quantification at all costs ignores the integrity of individuals and the qualitative differences between human values (Taylor, 1982). By defining “right” as that which maximizes a given “good” (Rawls, 1982, p. 183), it becomes blind to unintended consequences. As Oakley, Knafo, and McGrath (2012, p. 3) note, “some of the most horrific episodes in human history have resulted from well-intentioned altruistic tendencies”. This problem is illustrated by the story of a local mosquito net manufacturer whose business is destroyed by massive imports of free nets from the West. A short-term effective and measurable action thus leads to long-term systemic damage: the destruction of local industry and the deepening of dependence (Moyo, 2009, chap. 3).

EA advocates such as Yudkowsky (2015, p. 18) argue that, precisely because intuition fails, we must rely on cold calculation. For example, he cites an experiment in which people were willing to pay almost the same amount to save 2,000, 20,000, or 200,000 birds, influenced by an emotional image. However, this call for pure rationality is misleading and ignores at least

seven key aspects:

1. Value pluralism: The assumption that “number of lives saved” is the only parameter ignores dignity, autonomy, and ecological balance.
2. Systemic costs: The calculation ignores unintended consequences, as in the case of mosquito nets.
3. Motivational suppression: An emphasis on calculation can suppress “warm-glow giving”, which is a powerful driver of prosocial behavior (Elster, 2011).
4. Cognitive limits: The calculation itself is prone to bias, and the feeling of certainty is not a product of reason but a primary brain mechanism (Burton, 2012, p. 132).
5. Aggregation problem: The very act of adding up utility is philosophically controversial.
6. Agenda and power: The choice of what to measure is a political act that reinforces the power of those who define the metrics.
7. Epistemic arrogance: The claim that we have enough data to make reliable decisions is a manifestation of arrogance in the face of radical uncertainty.

This reductionism manifests in claims such as Alexander’s (2015, p. 28) assertion that “there is only one best charity: the one that helps most people the greatest amount per dollar”. Similarly, Singer (2015, pp. 119–120) argues that in the world we live in, donating to opera houses and museums is unlikely to do the best we can. While some EA-aligned philosophers have acknowledged that impartial maximization faces legitimate challenges from competing moral considerations (Pummer, 2019; Greaves, 2019), the dominant practice of organizations like GiveWell continues to privilege simple, quantifiable metrics over value pluralism. Before EA dismisses art as a luxury, it should assess its potential impact. Wouldn’t creating art that leads to the end of war be an incomparably more effective way of saving lives? Supporting art can be an investment in mental health, prevention, or the formation of moral imagination. To reject these possibilities *a priori* is to capitulate to complexity in favor of seductively simple but misleading mathematics.

The abstract nature of morality and the disregard for human embeddedness

The third problem lies in the abstract and decontextualized nature of EA’s moral principles. These are often presented as universal truths, independent of specific social and political practices. This recalls the difference between “ethical” and “political” concepts of rights, where the latter are “practice-dependent” and require interpretation of the real world. The utilitarian principle of EA, on the other hand, is a model of practice-independent ethics that attempts to apply an abstract formula to complex reality.

This abstractness ignores that morality is not only about maximizing good, but also about the ability to “justify one’s actions to others on grounds that they could not reasonably reject” (Scanlon, 1982, p. 116). Moral action is a social act, a dialogue, not a monological calculation. Instead of searching for a single “correct” answer through calculation, emphasis should be placed on developing the ability to engage in rational dialogue and consider different perspectives (Jedličková, 2025). The idea of a purely rational agent, which utilitarianism presupposes, is a fiction, because human desires and goals are always “permeated by memories and local ties and historical associations” (Hampshire, 1982, p. 154; Traphagan, 2012, p. 272). As Foster and Herring (2015, pp. 44–45) emphasize, we do not begin our moral lives as free, isolated subjects who subsequently choose commitments. On the contrary, “we begin our lives in relationships... and with relationships comes responsibility”. EA’s effort to create a universal, impartial, and practice-independent system thus appears as another attempt in a long line of modern moral philosophy. Unconditional principles such as the maximization of good or Kant’s categorical imperative can be seen as a “search for new sources of normativity” in a post-religious world (Korsgaard, 1996, p. 18) or as systems that fulfill the function of secularized religion.

However, the EA approach ignores the fact that morality is not formed in an abstract space, but in specific communities. A necessary prerequisite for morality is loyalty to a specific community and its relationships. What subjects adopt as a norm is never morality as such, but always the specific morality of a particular social order, justified in terms of the forms of good applied in the life of that community. EA principles may therefore be acceptable within their community, but that does not mean that others should identify with them or that they should be elevated to a universal standard.

What and how we consider to be help is therefore deeply culturally conditioned, as shown by differences in volunteering and informal help across Europe (Glanville, Paxton, & Wang, 2016, cited in Herzog, 2020, p. 88) or different motivations for giving shaped by religion (Warner et al., 2015). EA's attempt to introduce a single, rational yardstick can thus become a form of moral imperialism under the guise of universality. When GiveWell applies its "moral weights" derived from a handful of survey questions in two African countries as a global standard for aid allocation, it does not show respect for local values but imposes a utilitarian, Western model of ethical decision-making on the whole world, ignoring the richness and diversity of human moral traditions. Ironically, the very study intended to provide a unified voice for "recipients" revealed the opposite. It showed that there is no universal "beneficiary preference", but rather big cultural differences. While in Kenya, only 24% of respondents refused to do so, in neighboring Ghana, it was more than half (52%). The very attempt to find a single, aggregated answer that could be used as a global standard thus failed on its own data (IDinsight, 2019). Instead of confirming universality, it revealed particularities that the EA system, in its quest for global effectiveness, must inevitably smooth over and ignore.

The model of a human being as a "billiard ball in a City suit" (Foster & Herring, 2015, p. 1), i.e., an atomistic and calculating unit, is in direct contradiction to our knowledge of human nature (Švec, 2020). Altruism and cooperation are driven by "emotional commitments, concern for the welfare of others, and a sense of moral obligation" (Simpson & Willer, 2015, cited in Herzog, 2020, p. 26). Our moral behavior is based on biologically rooted mechanisms such as mirror neurons, which "blur the boundaries between us and others" (Foster & Herring, 2015, p. 25), or nervous systems originally designed for caring for offspring. This is further supported by the concept of *prehensile compassion* – a unique human instinct to reduce suffering that has evolved beyond mere kin selection to a generalized application (Braus, 2025). Our capacity for empathy is partly genetically determined (Knafo & Uzefovsky, 2015), and our prosocial personality has many components that go far beyond rational calculation (Knafo-Noam et al., 2015). We are beings whose consciousness is guided by the need for "social influence and relational value" (Henriques, 2020, p. 223), where the boundary between *I* and *you* is crossed at the level of desires themselves (Graham, 2004, pp. 56–57). This is confirmed by other studies (Andreoni & Rao, 2011; Herzog, 2020, p. 71; De Waal, 2011, pp. 138–139; Illingworth, 2011, p. 203).

The contradiction culminates in Singer's famous example of a drowning child (Singer, 2011, p. 17). The problem is not only that this is a misleading analogy that ignores the causal complexity of real help (Wenar, 2011, p. 105). This example works on two levels: on the one hand, it is a powerful emotional appeal that evokes guilt in the reader. On the other hand, as soon as the reader wants to question this emotional response and point to the importance of relationships or context, the argument is immediately withdrawn to the seemingly impenetrable ground of abstract logic and consistency. In reality, EA does not reject emotions; rather, it instrumentalizes them. Singer's example serves as emotional "bait" designed to trigger primary empathy in the donor. However, once this impulse is activated, EA performs an argumentative switch: it dismisses the emotion as unreliable and replaces it with an algorithm. This process dehumanizes altruism by utilizing compassion merely as motivational fuel for a system that no longer recognizes compassion as a legitimate reason in the final decision-making process.

This strategy can lead to the donor's real motivation in the EA system not being a rational effort to maximize good, but an irrational effort to alleviate their own guilt or experience the *warm glow of giving* (Elster, 2011, pp. 67–68). Moral decision-making becomes a tool for managing one's own feelings, which is in stark contrast to the proclaimed ideal of pure impartiality (Vinding, 2018, p. 7). Moreover, this ideal is problematic in itself because it ignores the moral importance of our personal relationships, which are not just psychological weaknesses but a defensible part of moral life (Pummer, 2019, p. 122; Williams, 2011, pp. 67–69). Even “rational” calculations and algorithms themselves may be merely a reproduction of our existing prejudices (Tasioulas, 2019, p. 342).

The attempt to bypass these prejudices through the perspective of *the universe* (Singer & de Lazari-Radek, 2014) assumes that objectivity is proportional to abstraction. Yet this overlooks the necessity of engaged rationality—the idea that certain moral truths are accessible only through our relational situatedness. A relationship is not a “bias” but an epistemic tool. Furthermore, the EA framework ignores the issue of moral feasibility: as Law, Campbell, and Gaesser (2021) demonstrate, impartial helping often undermines the very structures of social reciprocity it requires. Singer and de Lazari-Radek (2017) attempt to solve this by suggesting that utilitarianism can be self-effacing, meaning it is sometimes rational to ignore its calculations in practice. This, however, creates a fatal paradox: if EA must deny its own methodology to remain effective, it ceases to be a coherent guide for action. By forcing a logic that is fundamentally at odds with human nature into the realm of totalizing moral obligation, the system collapses under its own weight.

Ultimately, this transfers an unbearable responsibility onto the individual. If an effective altruist is truly responsible for maximizing global good, then they must be accountable for every decision they make in light of infinite possibilities. What if they are wrong? What if, by helping one organization, they neglect to support another that cares for a neglected individual who, without help, would become “Hitler 10.0”, a tyrant who uses AI to enslave humanity? How can an effective altruist be sure that his seemingly simple and rational calculation is not in fact a manifestation of extreme moral irresponsibility? If they are wrong, can the victims of their miscalculation make claims against them? Or will they simply apologize with the words, “Sorry, we didn’t have enough evidence?” The very requirement for such a calculation leads either to paralysis or to the use of defense mechanisms that protect us from unacceptable thoughts (Allen, 2020, p. 67).

The problem extends beyond the psychological burden on the individual to systemic and political consequences. Utilitarianism, which focuses on isolated effective actions, overlooks the fact that their uncoordinated sum can lead to structural injustice. As De Wispelaere (2004, p. 17) points out, the result can be an ethically suboptimal distribution in which “some poor people receive more gifts, while others receive none”. This approach, which glorifies the autonomous donor, paradoxically leads to a diffusion of real responsibility for systemic outcomes (Clapham, 2007, p. 21). Ultimately, it frees philanthropists from responsibility for the unintended but harmful consequences of their decisions.

Criticism deepens further at the political level, where philanthropy replaces justice. This depoliticization operates both in principle and in practice, and the two are inseparable. In principle, EA's utilitarian framework treats existing economic structures not as objects of moral evaluation requiring political transformation, but as morally neutral constraints on optimization—tools to be exploited for maximum donation capacity. The focus on individual calculation and resource allocation systematically deflects attention from collective political action and structural critique. In practice, this manifests in EA's dominant career advice and giving patterns, which actively discourage engagement with systemic change.

The extreme demands of utilitarianism, which require individuals to donate until marginal utility is reached, shift responsibility for systemic problems such as poverty from the political

level to individual donors (de Lazari-Radek & Singer, 2017, pp. 75–79). The process of depoliticization culminates in the concept of *Earning to Give*. MacAskill (2015a, pp. 67–68; Vinding, 2018, p. 26) explicitly argues that an altruist should consider becoming “a highly paid banker and donate 50% of one’s income rather than work in the nonprofit sector”, presenting this as straightforwardly better for the world because the quantifiable donations outweigh other considerations. This argument presents the existing economic system as a neutral tool for generating philanthropic resources. Instead of questioning whether this system causes inequality, it legitimizes it as a source of funds for charity, making EA not only apolitical but also a system-preserving approach. Under the guise of a “global civil society”, EA promotes a specific, power- and ideology-driven agenda that bypasses democratic processes and sovereignty in favor of global governance by a narrow group of experts and philanthropists (Anderson, 2011, pp. 153, 163; Reich, 2011, pp. 182–185).

Consider a concrete example of this optimization logic clashing with systemic reform: An individual following EA principle might choose to work in the financial sector—perhaps in proprietary trading or hedge fund management—earning \$500,000 annually and donate \$250,000 to highly effective global health charities. According to EA calculations, these donations could save approximately 70 lives per year through malaria prevention. This is highly measurable and demonstrably effective. However, the alternative path—working as a financial regulation advocate or policy researcher at a fraction of the salary—might contribute to regulatory reforms that prevent future financial crises. Research by Maruthappu et al. (2016) demonstrates that the 2008 financial crisis was associated with approximately 260,000 excess cancer deaths in OECD countries alone during 2008–2010, primarily due to unemployment-related loss of health coverage and delayed diagnoses. If regulatory advocacy work contributed marginally to preventing or mitigating one future crisis of similar scale, its impact would dwarf the cumulative effect of decades of individual charitable donations. Yet the EA framework systematically undervalues this path because the impact is diffuse, temporally distant, and difficult to attribute to individual effort—precisely the characteristics of systemic political change. The optimization logic thereby becomes a barrier to addressing root causes, favoring quantifiable individual interventions over collective action against structural harm. Ultimately, the insistence on donation as the primary moral response ignores the fact that providing assistance often depends on an agent’s resources, skills, and personal integrity, rather than on a universal duty to maximize ROI (Luczak, 2025).

EA advocates might argue that this criticism is outdated because the modern movement is now also concerned with systemic change or so-called longtermism—a focus on the long-term future of humanity and existential risks. However, this shift does not solve the fundamental problem; on the contrary, it takes it to a logical extreme. If quantifying the impact of saving lives from malaria is fraught with uncertainty and value judgments, how much more speculative is the attempt to calculate the “expected value” of preventing the risks of AI 200 years from now?

Consider a concrete example: long-termists might argue that AI research should be slowed or regulated because there is a non-zero probability of an existential catastrophe within the next 200 years. Suppose their calculations yield an expected utility of slowing AI research as $E_1 = p \times U$, where p is the estimated probability of catastrophe (say, 5%) and U is the negative utility of human extinction (say, -10^{23} future lives lost). The alternative—continuing at the current pace—has expected utility E_2 . However, *Knightian uncertainty* means we do not know the probability p (it is speculative), nor the parameter U (how many people would exist in the future?), nor can we foresee the indirect consequences of slowing AI (lost medical breakthroughs? geopolitical instabilities?). Equally, we can construct a contradictory scenario: rapid AI development might solve the climate crisis and prevent civilization collapse, thereby saving an identical number of future lives. Under *Knightian uncertainty*, expected value

calculation thus collapses into an arbitrary choice between equally unsubstantiated narratives, where each parameter can be replaced with an estimate of several orders of magnitude different, with no estimate having empirical grounding. The *tyranny of the measurable* is replaced by the tyranny of the unimaginable, where even greater moral decisions are justified by even more fragile and untestable calculations.

This transition from empirically grounded philanthropy to speculative ideology directly impacts individual moral reasoning. In the context of our opening dilemma, the graduate's decision-making becomes even more absurd: she no longer chooses merely between a certain income at a bank and caring for her mother but is urged by the movement to seek influence within the US AI ecosystem based on the vague promise of altering the trajectory of artificial intelligence. Yet, while abandoning her mother is presented as a rational toll for "saving the future", the movement's own current manuals admit that an individual's real impact in this field is "lumpy, hard to predict" and contingent on uncontrollable "policy windows" (80,000 Hours, 2025). If a bank donation was at least a tangible (albeit reduced) contribution, influence over an "AI Takeover" remains entirely within the sphere of total *Knightian uncertainty*. EA thus positions the daughter as a moral gambler: she is asked to sacrifice the certain good of a family relationship for a statistically ungrounded chance that she might find herself "in the right room" at the moment of a random political crisis.

This shift toward longtermism does not eliminate EA's methodological arrogance but rather radicalizes it. While malaria nets present us with the *tyranny of the measurable*, calculations of existential AI risks lead us into the tyranny of the unimaginable. Here, scientific objectivity dissolves into pure mathematical fiction, where minute changes in speculative probabilities lead to astronomical differences in expected value. Consequently, EA ceases to be an empirically grounded philanthropy and becomes an ideology that, in the name of hypothetical billions of future humans, legitimizes the neglect of present, tangible suffering.

Instead of reductionist attempts at quantification, we should look for concepts such as *moral literacy* (Herman, 2007, p. 80) – the learned ability to interpret moral situations in their complexity – or "generosity of attention and emotion" (Smith & Davidson, 2014, p. 18; Foster & Herring, 2015, p. 14). Altruism is not the solution to a mathematical problem. It is a complex human practice deeply rooted in our biology, emotions, and social relationships. The key question remains whether effective altruism, grounded in utilitarianism, can integrate these dimensions without contradicting itself (de Lazari-Radek & Singer, 2017, p. 35). Attempts to reformulate the model are met with pessimism as to whether a model whose very essence is calculation can ever truly reconcile itself with the incalculable.

Calculated philanthropy versus selfless altruism

Ultimately, criticism of the reductionist EA model leads us to a fundamental conceptual question: Is what we are talking about really altruism? This distinction does not necessarily rest on the degree of self-sacrifice—as Singer (1972, 2015) rightly notes, a moral agent is not required to sacrifice their own well-being to the point of comparable moral loss. However, the problem lies in the ontological nature of the act and the mode of social participation. We must distinguish between altruism as an ethical relationship and "effective philanthropy" as a technocratic institution.

This distinction is fundamental to our critique: altruism is essentially an interpersonal relationship—a direct moral encounter between *I and Thou*—whereas effective philanthropy is an administrative relationship to a portfolio, a cold calculation between *I and Numbers*. While altruism, in its classical sense, is a direct moral response to the suffering of the Other rooted in empathy and recognition (Monroe, 1996), EA transforms this impulse into a systematic "investment logic" (Synowiec, 2016). Within the EA framework, help ceases to be a relational act and becomes a form of "meta-charity"—a strategic evaluation of causes from the outside,

in which the donor acts as a portfolio manager of social impact (Synowiec, 2016, pp. 151–152). This shift is not merely a refinement of altruistic practice; it is a transition to a different category of social action.

In this system, the recipient is no longer a specific person with a face and dignity, but an “investment opportunity” with a projected Return on Investment (ROI). This “universal mathematics of morality” assumes that a 90% chance of helping 100 people is strictly equivalent to a 100% chance of helping 90 people, manifesting a reified mindset that instrumentalizes compassion to maintain a mechanistic worldview. Therefore, the name “effective altruism” is misleading. It is effective philanthropy: a sophisticated, data-driven system for allocating surplus resources to maximize measurable benefits while preserving the donor’s fundamental interests and the existing economic order. Recognizing this difference is not a semantic exercise but an analytical necessity to understand what is lost in the pursuit of efficiency: the immediate ethical responsibility that defies quantification.

Conclusion

The path of EA, which began with a seemingly simple choice between saving anonymous children and caring for one’s own mother, has led us to profound and disturbing questions about the nature of morality, rationality, and humanity itself. As this analysis has shown, EA’s attempt to turn altruism into an exact science based on utilitarian calculation fails on three fundamental levels: in its mistaken understanding of the human subject as a rational utility maximizer; in the *tyranny of the measurable*, which systematically ignores unquantifiable values; and in its promotion of an abstract, decontextualized morality that overlooks the social and relational anchoring of human beings.

The ambition to rationalize aid is not in itself reprehensible. There is room for more systematic reflection on the consequences of our actions. The problem arises, however, when this tool is elevated to a universal moral compass and its methodological fragility is masked by a facade of scientific objectivity. The example of the GiveWell organization and its “moral weights” has exposed this process not as a search for truth, but as an act of performative rationality—a sophisticated exercise that legitimizes the authority of the donor and their culturally conditioned prejudices, while voices of dissent and the incomparability of human lives are coldly incorporated as additional data points into a pre-given model. The greatest threat to EA is thus not a lack of data, but epistemic arrogance: the belief that the complex social world can be reduced to a set of optimization problems. This immodesty not only leads to dubious conclusions but, in the long run, undermines the very public trust in scientific (rational) authority that it seeks to rely on.

Ultimately, it seems that EA will have to face a much more fundamental choice: to abandon its utilitarian core. This is not just a philosophical dispute with alternative ethical theories, but a practical necessity. As the gap between proclaimed ideals and actual human behavior—even among the movement’s own advocates—shows, utilitarian maximalism is unsustainable for social and emotional beings such as us. Its strict demands create an elitist moral code for a narrow group of insiders, but fail as a guide for everyday moral life, which is inextricably linked to relationships, loyalty, and context. Rational moralizing, detached from this lived reality, becomes sterile.

However, if the movement is to overcome these limitations and fulfill its original promise of doing good, it could take three constructive paths that directly address each identified failure:

1. From quantitative hegemony to methodological pluralism (addressing preference instability): Since people lack stable, well-defined preferences even about basic choices, EA’s reliance on preference maximization becomes meaningless. Instead of fetishizing randomized controlled trials and metrics such as QALYs, EA could integrate qualitative methods, ethnographic studies, and historical analysis that acknowledge the provisional and contextual

nature of human values. This would allow for an understanding of the systemic causes of suffering and an appreciation of immeasurable values such as dignity, social capital, and cultural integrity without requiring impossible precision about unstable preferences.

2. From abstract calculation to contextual support (addressing adaptive preferences): Since preferences can be shaped by oppression and constrained circumstances, imposing external “moral weights” risks legitimizing unjust *status quos* by treating resigned acceptance as genuine satisfaction. Instead, attention should shift to empowering local communities to identify their own authentic needs and solutions, free from oppressive constraints. Effectiveness would thus be understood not as an external audit based on predetermined metrics, but as support for autonomy and long-term sustainability that enables communities to develop non-adaptive preferences.

3. From depoliticized charity to engaged justice (addressing relational partiality): Rather than dismissing people’s commitments to family, community, and place as cognitive biases that undermine global optimization, EA should recognize these relationships as legitimate moral starting points. The movement must abandon its systematic apoliticism and, instead of legitimizing inequalities through concepts like *earning to give*, actively examine the political and economic structures that cause these inequalities and support systemic change rooted in place-based commitments, even if its impact is not immediately or easily measurable. Ultimately, Singer’s expansion of this calculus to “astronomical” numbers of future lives (de Lazari-Radek & Singer, 2017, p. 112) represents the final stage of this de-realization. By prioritizing the *Total View*—where a massive population of barely positive lives is superior to a smaller, thriving one (ibid, p. 114)—EA moves from compassionate aid to a form of cosmic accounting. This “calculated philanthropy” manages a portfolio of lives while bypassing the political structures of guilt and responsibility.

The term “effective altruism” is ultimately misleading. What we have analyzed is rather effective philanthropy—a perfectly optimized, yet ultimately dehumanized, calculation of resource allocation. True altruism, embodied in the selfless act of a rescuer during the Holocaust, defies calculation and stems from a deep conviction about the interconnectedness of human destinies. The question remains whether a project that began with the pursuit of effective altruism is willing, in its quest for efficiency, to accept the incalculable, or whether, in the name of rationality, it will definitively sacrifice the very essence of human solidarity.

Acknowledgement

This study was prepared as part of the VEGA project No. 2/0118/24, *The Social Nature of Normativity as a Starting Point for Explaining the Relationship between Individual and Collective Autonomy*.

References

- 80,000 HOURS (2025): The US AI policy landscape: where to work to have the biggest impact. In: *80,000 Hours* [online] [Retrieved February 7, 2026] Available at: <https://80000hours.org/articles/the-us-ai-policy-landscape-where-to-have-the-biggest-impact/>.
- ALEXANDER, S. (2015): Efficient charity: Do unto others. In: R. Carey (ed.): *The Effective Altruism Handbook*. Oxford: Centre for Effective Altruism, pp. 25–31.
- ALLEN, D. M. (2020): Mental and interpersonal mechanisms of groupthink maintenance. In: D. M. Allen & J. W. Howell (eds.): *Groupthink in Science: Greed, pathological altruism, ideology, competition and culture*. Springer, pp. 54–78.
- ANDERSON, K. (2011): Global philanthropy and global governance: The problematic moral legitimacy of the relationship between global civil society and the United Nations. In: P. Illingworth, T. Pogge & L. Wenar (eds.): *Giving Well: The ethics of philanthropy*. Oxford: Oxford University Press, pp. 149–176.

- ANDREONI, J. & RAO, J. M. (2011): The power of asking: How communication affects selfishness, empathy, and altruism. In: *Journal of Public Economics*, 95(7-8), pp. 513–520.
- BANERJEE, A. V. & DUFLO, E. (2019): *Good economics for hard times*. New York: PublicAffairs.
- BARANOVA, S. (2025): Infanticide in bioethics: Singer's argument and cross-framework debates. In: *Ethics & Bioethics (in Central Europe)*, 15(3–4), pp. 151–165. doi:10.2478/ebce-2025-0017.
- BATÔT, L. & BELLI, A. (2025): Algorithmic care: Peter Singer's ethics and the challenge of AI in the end-of-life medicine. In: *Ethics & Bioethics (in Central Europe)*, 15(3–4), pp. 192–203. doi:10.2478/ebce-2025-0018.
- BRAUS, A. J. (2025): Moral status, suffering, and compassion: Towards reconciling human moral priority with animal welfare. In: *Ethics & Bioethics (in Central Europe)*, 15(3–4), pp. 166–176. doi:10.2478/ebce-2025-0011.
- BURTON, R. A. (2012): Pathological certitude. In: B. Oakley, A. Knafo, G. Madhavan & D. S. Wilson (eds.): *Pathological altruism*. New York: Oxford University Press, pp. 131–137.
- CLAPHAM, A. (2007): *Human rights: A very short introduction*. Oxford: Oxford University Press.
- CRISP, R. (2025): The objectivity of ethics and the unity of practical reason. In: *Ethics & Bioethics (in Central Europe)*, 15(3–4), pp. 204–208. doi:10.2478/ebce-2025-0019.
- DE LAZARI-RADEK, K. & SINGER, P. (2014): *The point of view of the Universe: Sidgwick and contemporary ethics*. Oxford: Oxford University Press.
- DE LAZARI-RADEK, K. & SINGER, P. (2017): *Utilitarianism: A very short introduction*. Oxford: Oxford University Press.
- DE WAAL, A. (2011): Ethics in translation: Principles and power in the philanthropic encounter. In: P. Illingworth, T. Pogge & L. Wenar (eds.): *Giving Well: The ethics of philanthropy*. Oxford: Oxford University Press, pp. 133–148.
- DE WISPELAERE, J. (2004): Altruism, impartiality and moral demands. In: J. Seglow (ed.): *The ethics of altruism*. London: Frank Cass, pp. 9–33.
- DIMONYE, S. C., AYOZIE, J. N. & OKEKE, N. M. (2025): A critique of Peter Singer's utilitarian perspective on abortion: An African (Igbo) jurisprudence approach. In: *Ethics & Bioethics (in Central Europe)*, 15(3–4), pp. 177–191. doi:10.2478/ebce-2025-0016.
- ELSTER, J. (1982): Sour grapes – utilitarianism and the genesis of wants. In: A. Sen & B. Williams (eds.): *Utilitarianism and beyond*. Cambridge: Cambridge University Press, pp. 219–238.
- ELSTER, J. (2011): The Valmont effect: The Warm-Glow theory of philanthropy. In: P. Illingworth, T. Pogge & L. Wenar (eds.): *Giving Well: The ethics of philanthropy*. New York: Oxford University Press, pp. 67–83.
- FOSTER, C. & HERRING, J. (2015): *Altruism, welfare and the law*. Springer.
- GIVEWELL (2015): Your dollar goes further overseas. In: R. Carey (ed.): *The effective altruism handbook*. Oxford: Centre for Effective Altruism, pp. 83–84.
- GIVEWELL (2019): Research on moral weights – 2019. In: GiveWell [online] [Retrieved February 2, 2026] Available at: <https://www.givewell.org/how-we-work/our-criteria/cost-effectiveness/2019-moral-weights-research>
- GIVEWELL (2025): Our top Charities. In: GiveWell [online] [Retrieved September 28, 2025]. Available at: <https://www.givewell.org/charities/top-charities>
- GLANVILLE, J. L., PAXTON, P. & WANG, Y. (2016): Social capital and generosity: A multilevel analysis. In: *Nonprofit and Voluntary Sector Quarterly*, 45(3), pp. 526–547.
- GLUCHMAN, V. (2017): G. E. Moore and theory of moral/right action in ethics of social consequences. In: *Ethics & Bioethics (in Central Europe)*, 7(1–2), pp. 57–65. doi:10.1515/ebce-2017-0002.

- GRAHAM, K. (2004): Altruism, self-interest and the indistinctness of Persons. In: J. Seglow (ed.): *The ethics of altruism*. London: Frank Cass, pp. 51–70.
- GREAVES, H. (2019): Population and life extension. In: D. Edmonds (ed.): *Ethics and the contemporary world*. London & New York: Routledge, pp. 43–54.
- GREEN, D. (2008): *From poverty to power: How active citizens and effective states can change the world*. Oxford: Oxfam International.
- HAMMOND, P. J. (1982): Utilitarianism, uncertainty and information. In: A. K. Sen & B. Williams (eds.): *Utilitarianism and beyond*. Cambridge: Cambridge University Press, pp. 85–102.
- HAMPSHIRE, S. (1982): Morality and convention. In: A. Sen & B. Williams (eds.): *Utilitarianism and beyond*. Cambridge: Cambridge University Press, pp. 145–157.
- HENRIQUEZ, D. C. (2016): *Effective altruism and do no harm: A case for harm-sensitive effective altruism*. Saarbrücken: Scholar's Press.
- HENRIQUES, G. (2020): Groupthink and the evolution of reason-giving. In: D. M. Allen & J. W. Howell (eds.): *Groupthink in science: Greed, pathological altruism, ideology, competition and culture*. Cham: Springer, pp. 184–223.
- HERMAN, B. (2007): *Moral literacy*. Cambridge, MA: Harvard University Press.
- HERZOG, P. S. (2020): *The science of generosity: Causes, manifestations, and consequences*. Cham: Palgrave Macmillan.
- IDINSIGHT (2019): Beneficiary Preferences: Findings from Ghana and Kenya. In: *IDinsight* [online] [Retrieved July 28, 2026]. Available at: <https://www.idinsight.org/reports-and-publications/beneficiary-preferences/>
- ILLINGWORTH, P. (2011): Giving back: Norms, ethics, and law in the service of philanthropy. In: P. Illingworth, T. Pogge & L. Wenar (eds.): *Giving Well: The ethics of philanthropy*. New York: Oxford University Press, pp. 196–217.
- JEDLIČKOVÁ, A. (2025): Ethical decision-making in academic contexts: The impact of the KMDD® approach on moral competence. In: *Ethics & Bioethics (in Central Europe)*, 15(1–2), pp. 120–135.
- KALAJTZIDIS, J. (2025): Debt, interest, and effective altruism: Critically evaluating Singer's utilitarian ethics in the context of contemporary economics. In: *Ethics & Bioethics (in Central Europe)*, 15(3–4), pp. 233–245. doi:10.2478/ebce-2025-0014.
- KARASTERGIOU, A. (2025): Data Ethics: Issues related to data biases and the application of traditional ethical theories on AI Ethics. In: *Ethics & Bioethics (in Central Europe)*, 15(1–2), pp. 107–119.
- KISSEL, J. (2017): Effective altruism and anti-capitalism: An attempt at reconciliation. In: *Essays in Philosophy*, 18(1), pp. 1–31. doi:10.7710/1526-0569.1573.
- KNAFO, A. & UZEFOVSKY, F. (2015): Variation in empathy: The interplay of genetic and environmental factors. In: D. W. Haley & M. H. Bornstein (eds.): *The infant mind: Origins of the social brain*. New York: Guilford Press, pp. 195–214.
- KNAFO-NOAM, A., UZEFOVSKY, F., ISRAEL, S., DAVIDOV, M. & ZAHN-WAXLER, C. (2015): The prosocial personality and its facets: Genetic and environmental architecture of mother-reported behavior of 7-year old twins. In: *Frontiers in Psychology*, 6, 112, pp. 1–9.
- KORSGAARD, C. M. (1996): *The sources of normativity*. Cambridge: Cambridge University Press.
- LAW, K. F., CAMPBELL, D. & GAESSER, B. (2021): Biased benevolence: The perceived morality of effective altruism across social distance. In: *Personality and Social Psychology Bulletin*, 47(8), pp. 1263–1279. <https://doi.org/10.1177/01461672211002773>
- LIU, S. (2025): Beyond efficiency: Reconciling Singer's effective altruism with care ethics. In: *Ethics & Bioethics (in Central Europe)*, 15(3–4), pp. 209–219. doi:10.2478/ebce-2025-0013.
- LUCZAK, J. (2025): Rethinking Singer: Toward a valid argument for helping the global poor.

- In: *Ethics & Bioethics (in Central Europe)*, 15(3–4), pp. 220–232. doi:10.2478/ebce-2025-0012.
- MACASKILL, W. (2015a): Don't get a job at a charity: Work on Wall Street. In: R. Carey (ed.): *The effective altruism handbook*. Oxford: Centre for Effective Altruism, pp. 67–70.
- MACASKILL, W. (2015b): What is effective altruism? In: R. Carey (ed.): *The effective altruism handbook*. Oxford: Centre for Effective Altruism, pp. 11–13.
- MARUTHAPPU, M., WATKINS, J., NOOR, A. M., WILLIAMS, C., ALI, R., SULLIVAN, R., ZELTNER, T. & ATUN, R. (2016): Economic downturns, universal health coverage, and cancer mortality in high-income and middle-income countries, 1990–2010: a longitudinal analysis. In: *The Lancet*, 388, pp. 684–695.
- MIRRELES, J. A. (1982): The economic uses of utilitarianism. In: A. Sen & B. Williams (eds.): *Utilitarianism and beyond*. Cambridge: Cambridge University Press, pp. 63–84.
- MONROE, K. R. (1996): *The heart of altruism: Perceptions of a common humanity*. Princeton, NJ: Princeton University Press.
- MOYO, D. (2009): *Dead Aid: Why Aid Is Not Working and How There Is a Better Way for Africa*. New York: Farrar, Straus and Giroux.
- NG, Y-K. (2020): Effective Altruism despite the second-best challenge: Should Indirect Effects Be Taken into Account for Policies for a Better Future? In: *Futures*, 121, pp. 1–10. doi:10.1016/j.futures.2020.102568.
- OAKLEY, B., KNAFO, A. & MCGRATH, M. (2012): Pathological altruism—an introduction. In: B. Oakley, A. Knafo, G. Madhavan & D. S. Wilson (eds.): *Pathological Altruism*. New York: Oxford University Press, pp. 3–9.
- PUMMER, T. (2019): Charity and partiality. In: D. Edmonds (ed.): *Ethics and the Contemporary World*. London & New York: Routledge, pp. 121–132.
- RAWLS, J. (1982): Social unity and primary goods. In: A. Sen & B. Williams (eds.): *Utilitarianism and beyond*. Cambridge: Cambridge University Press, pp. 159–185.
- REICH, R. (2011): Toward a political theory of philanthropy. In: P. Illingworth, T. Pogge, & L. Wenar (eds.): *Giving Well: The Ethics of Philanthropy*. Oxford: Oxford University Press, pp. 177–195.
- SCANLON, T. M. (1982): Contractualism and utilitarianism. In: A. Sen & B. Williams (eds.): *Utilitarianism and beyond*. Cambridge: Cambridge University Press, pp. 103–129.
- SCHWEIGER, G. & GRAF, G. (2015): *A Philosophical Examination of Social Justice and Child Poverty*. Basingstoke: Palgrave Macmillan.
- SEN, A. & WILLIAMS, B. (1982): Introduction: Utilitarianism and beyond. In: A. K. Sen & B. Williams (eds.): *Utilitarianism and beyond*. Cambridge: Cambridge University Press, pp. 1–21.
- SHAW, C. & VANADIA, S. (2022): Utilitarianism on the front lines: COVID-19, public ethics, and the “hidden assumption” problem. In: *Ethics & Bioethics (in Central Europe)*, 12(1–2), pp. 60–78.
- SIMPSON, B. & WILLER, R. (2015): Beyond altruism: Sociological foundations of cooperation and prosocial behavior. In: *Annual Review of Sociology*, 41(1), pp. 43–63.
- SINGER, P. (1972): Famine, affluence, and morality. In: *Philosophy & Public Affairs*, 1(3), pp. 229–243.
- SINGER, P. (2011): What should a billionaire give—and what should you? In: P. Illingworth, T. Pogge & L. Wenar (eds.): *Giving Well: The ethics of philanthropy*. New York: Oxford University Press, pp. 13–25.
- SINGER, P. (2015): *The most good you can do: How Effective Altruism is changing ideas about living ethically*. New Haven: Yale University Press.
- SMITH, C. & DAVIDSON, H. (2014): *The paradox of generosity: Giving we receive, grasping we lose*. Oxford: Oxford University Press.

- SYNOWIEC, J. (2016): Ethics for everyday heroes – from Utilitarianism to Effective Altruism. In: *Ethics & Bioethics (in Central Europe)*, 6(3–4), pp. 147–156. doi:10.1515/ebce-2016-0015.
- ŠVEC, J. (2020): K historickým zdrojom liberalizmu – medzi politickou ideológiou a filozofickými pojmami [On historical sources of liberalism – between political ideology and philosophical concepts]. In: *Filozofia*, 75(5), pp. 341–355. doi:10.31577/filozofia.2020.75.5.1
- TASIOULAS, J. (2019): AI and robot ethics. In: D. Edmonds (ed.): *Ethics and the Contemporary World*. London & New York: Routledge, pp. 335–352.
- TAYLOR, C. (1982): The diversity of goods. In: A. Sen & B. Williams (eds.): *Utilitarianism and beyond*. Cambridge: Cambridge University Press, pp. 129–144.
- TRAPHAGAN, J. W. (2012): Altruism, pathology, and culture. In: B. Oakley, A. Knafo, G. Madhavan & D. S. Wilson (eds.): *Pathological Altruism*. New York: Oxford University Press, pp. 271–287.
- VINDING, M. (2018): *Effective Altruism: How can we best help others?* North Carolina: CreateSpace Independent Publishing Platform.
- WARNER, C. M., KILINÇ, R., HALE, C. W., COHEN, A. B. & JOHNSON, K. A. (2015): Religion and public goods provision: Experimental and interview evidence from Catholicism and Islam in Europe. In: *Comparative Politics*, 47(2), pp. 189–209.
- WENAR, L. (2011): Poverty Is No Pond: Challenges for the Affluent. In: P. Illingworth, T. Pogge & L. Wenar (eds.): *Giving Well: The Ethics of Philanthropy*. Oxford: Oxford University Press, pp. 104–132.
- WILLIAMS, B. (2011): *Ethics and the limits of philosophy*. London & New York: Routledge.
- YUDKOWSKY, E. (2015): Scope neglect. In: R. Carey (ed.): *The Effective Altruism Handbook*. Oxford: Centre for Effective Altruism, pp. 14–18.