

Doing Good Better: What We Can Learn from Effective Altruism

Alexander Herwix¹

¹ Cologne Institute for Information System, University of Cologne, Germany
herwix@wiso.uni-koeln.de

Abstract. Over the last few decades Information Systems (IS) research has been striving but also struggling to demonstrate relevance and real world impact. We provide a new ethical perspective on this important challenge by framing IS research in the terms of Effective Altruism (EA). EA is a growing global movement and research project focused on a deceptively simple-seeming question: How can we do the most good with the resources we have? This paper gives an overview of EA and investigates what IS research may learn from the growing body of knowledge emerging from EA. We present two main ideas – *shared goals, principles and measures*, and *cause neutrality and focus area selection* – which seem fruitful for IS research to consider.

Keywords: Impact of IS, Effective Altruism, Ethics of IS

1 Introduction

A self-conscious view of the relevance, direction, and impact of our work has accompanied information system (IS) research almost from its inception [1–4] and continues to this day [5]. Most prominently, Benbasat and Zmud [3] escalated these concerns into a comprehensive discussion about the very identity of IS research. Their reasoning was that a dominant design in the form of common core of constructs and relationships was necessary for IS research to mature and gain relevance (i.e., cognitive legitimacy) with important stakeholders (e.g., governing bodies, public organizations, etc.). After all, how convincing and relevant can IS research be when it cannot demonstrate a consensus around its core concepts? This stance sparked a controversial and still unresolved discussion about the nature and direction of IS research [4]. To this day, the discourse on this topic remains controversial with different perspectives coexisting and competing [5].

This paper strives to broaden the discourse by focusing on the moral dimension of IS research. Here, we follow in the footsteps of Walsham [6], who has already argued for the need to reframe the discourse about the nature and identity of IS research and to reconsider its values and vision in the form of a strong ethical agenda:

Architects want to build better buildings, medics want to help people live longer and healthier lives, engineers want to build more effective technological systems to

improve efficiency and artists want to stimulate our subtler senses with their work. IS scholars and practitioners should be concerned with how to use ICTs to help make a better world, where everybody has the opportunity and capability to use technologies to make better lives for themselves, their communities and the world in general. [6]

Most importantly, emphasizing the ethical agenda of IS research has the virtue of providing a common ground on which most scholars should be able to unite. No matter what topics, concepts, theories, or methods IS scholars investigate or make use of, we should all be able to agree on the desire to be relevant and have a positive impact on the world [6]. Thus, focusing on the ethical dimension and agenda of IS research provides new opportunities for reframing the debate on IS identity as well as IS impact and may help to move it closer towards resolution.

Towards this goal, we extend Walsham's initial work by starting to investigate the foundational question that an ethical agenda for IS research poses: *how can we most effectively improve the world with IS research?* As a first step, we turn to *Effective Altruism* (EA) [7] for an expert perspective on this topic. EA is a growing global movement and research program focused on the deceptively simple-seeming question: *How can we do the most good with the resources we have?* EA advocates for an open, scientific mindset when considering moral questions [8] and, since its inception, the EA movement has curated a rapidly growing body of knowledge focused on the key considerations that should be kept in mind when one tries to effectively improve the world [7–12]. Thus, we argue that EA is a key reference discipline that should be considered when an ethical agenda for IS research is discussed.

The rest of the paper is structured as follows. First, we provide a short overview of EA and its key concepts and considerations. Second, we frame IS research in terms of EA and discuss how it could be informed by two key ideas from EA. Third, we conclude with a short discussion and outlook.

2 A Short Overview of Effective Altruism

Historically, the EA movement can be traced back to the philosophical arguments of the philosopher Peter Singer [13], who observed that humans are often in a great position to help avoid unnecessary suffering (e.g., by giving to effective charities) but somehow neglect to do this (e.g., due to cognitive biases). For example, while most people who walk by a shallow pond and see a drowning child would hurry to its rescue even at great cost to themselves, far fewer people would give a comparable amount of money to effective charities who could save the equivalent of the drowning child in a developing country. According to Singer [13], there is no morally relevant difference between the two situations and, thus, we ought to be impartial [14] and help in both cases. Since then, more and more people have taken Peter Singer up on his challenge to give morality a bigger role in their lives and united around what was to become the central question of EA: *How can we do the most good with the resources we have?*

Today, EA is a rapidly developing, and global movement of diverse people driven by the desire to effectively improve the world [15]. EA has started to integrate research and considerations from philosophy with other disciplines, most prominently, economics to provide a sound foundation for this endeavor [7, 10, 12]. Due to the scope and complexity of this work, only a short overview of major considerations can be presented and discussed in this paper. We follow the EA Concepts website [12] in structuring this overview and, thus, focus on considerations emerging from general features of the world and how these impact ethical decision-making.

2.1 General Features of the World

EA highlights, amongst others, three general features of the world, which are deemed highly relevant to ethical decision-making, namely, *capacity to feel pleasure and pain*, *future considerations*, and *variation in cost-effectiveness* [12].

Capacity to Feel Pleasure and Pain. The creation of pleasure and avoidance of pain is generally considered to be an important value in all plausible *theories of value* that underpin systematic moral considerations [12]. Thus, it is highly relevant to understand and/or determine to what degree pleasure and pain can be experienced by beings that may be affected by one's actions. Three major positions can generally be discerned in the philosophic and scientific discourse, namely, *only humans feel pleasure and pain*, *only sufficiently advanced animals feel pleasure and pain*, or *the capacity to feel pleasure and pain is substrate-independent and may arise in non-animal subjects as well* [12]. For example, while most people hold that only sufficiently advanced animals feel pleasure and pain, some people argue, for instance, that sufficiently advanced artificial intelligence (e.g., in the form of computer simulations of people) could have similar experiences to biology-based beings [16]. It is obvious that one's position on this question may greatly alter the moral weight associated with a given action and, thus, EA generally argues for deliberate engagement with this challenge.

For example, recognizing the possibility that conscious artificial intelligence may be created accidentally and unwittingly exposed to large amounts of suffering would likely influence the development and research of IS which could conceivably exhibit this property. However, it is also important to note that the recognition of other beings' capacity to experience does not necessarily imply that all beings' experiences need to be of equal importance in one's considerations [12].

Future Considerations. The observable value of any action today is determined by what happens in the future [12]. For example, on the one hand, if a large meteor was going to hit earth and wipe out humanity in a few years' time that would considerably decrease the value of work done to combat climate change [12]. On the other hand, if one sees future people as morally relevant and considers that in a positive (post-human) future [17] most people who will have ever lived have not yet been born. This would imply that all actions which help to positively shape the future are of overwhelming importance [18]. More specifically, if one accepts these premises, it is plausible that the most important thing about actions today are their long-run consequences in the far

future (e.g., via indirect flow-through effects) [18, 19]. Thus, EA generally advocates for the importance of future considerations even on present-day ethical decision-making [18, 19]. Of particular interest in this context are potentially transformative technologies such as artificial intelligence, genetic engineering, atomically precise manufacturing and methods of interstellar space travel [12] which are all interlinked with IS and, thus, relevant fields of IS research. New technologies can not only radically improve the world but also create new risks that may even lead to human extinction [20, 21].

Variation in Cost-Effectiveness. According to the best estimates resulting from empirically grounded cost-effectiveness analyses [11, 22] there are significant variations in the cost-effectiveness of interventions focused on improving the world [23, 24]. For example, the most cost-effective intervention focused on reducing the suffering caused by HIV/AIDS (education of high-risk individuals) is estimated to be around 1,400 times cheaper than the least cost-effective intervention (surgical treatment for Kaposi's sarcoma). This means that for the cost of treating one Kaposi's sarcoma about 1,400 high-risk individuals can be prevented from contracting HIV/AIDS through the means of education [23].

Although such cost-effectiveness estimates are inherently uncertain due to the variety in influencing factors and overall complexity and entanglement of the modern world, the magnitude of observed differences in cost-effectiveness estimations suggests that significant variation in cost-effectiveness exists to a reasonable degree of confidence and is likely a general feature across focus areas [22–24]. Consequently, in the light of general constraints on available resources, EA argues for a moral imperative to use the resources available wisely and in a cost-effective manner [23].

2.2 Ethical Decision-Making

Ethical decision-making is the study of how to make the decisions with the information that is available based on what one values [12]. It can be separated into *idealized ethical decision-making*, focused on applying standard decision theory to ethical problems, and *practical ethical decision-making*, focused on pragmatic approaches to ethical decision-making that try to make idealized ethical decision-making more tractable and applicable in real world contexts [12].

Idealized Ethical Decision-Making. In idealized ethical decision-making the behavior of perfectly rational ethical agents is studied [12]. Due to the expansive body of knowledge in this field and the length restrictions of this paper, we simply summarize that idealized ethical decision-making consists of three general branches, namely, *epistemology* (how to form beliefs given available evidence?), *ethics* (what is valuable?), and *decision theory* (how to reach a decision given values and beliefs?) [12]. EA advocates for the study of idealized ethical decision-making to inform and guide practical ethical decision-making.

Practical Ethical Decision-Making. In practical ethical decision-making, decisions are generally structured in terms of *problems*, *interventions*, and *focus areas* [12]. A problem is something true about the world, which, if it stopped being true, would improve the world [12]. For example, “people dying from malaria” would be a problem. Interventions are attempts to solve or make progress on problems [12]. For example, “distribution of bednets” would be an intervention aimed at solving “people dying from malaria”. A focus area is a bundle of (inter)related problems that make up a broad field of inquiry [12]. For example, “global health” would be a focus area, which would include the problem “people dying from malaria” but also other related problems such as “people suffering from neglected tropical diseases”.

The benefit of distinguishing between problems, interventions, and focus areas is that it allows for a structured approach to identify the best opportunities for doing good. Given the vast range of opportunities out there and the limited amount of resources available, it is necessary to engage in some form of triage or prioritization. Assessing the relevance of focus areas allows for a rough mapping of the overall problem space in terms of likelihood of containing great opportunities for doing good, which, in turn, allows for the prioritization of more in-depth investigations and engagement. Importantly, EA strives to be cause neutral and is open to focus on any focus area as long as it is likely to lead to cost-effective ways to improve the world [14, 25].

To facilitate this assessment of focus areas, the *importance*, *tractability*, *neglectedness*, (ITN) framework has emerged as a useful tool within the EA movement [12]. The ITN framework holds that a focus area is more likely to contain great opportunities for doing good the more *important* (the more important the problem, the higher the payoff), *tractable* (the more tractable the problem, the less resources need to be invested), and *neglected* (the more neglected the problem, the higher is the expected marginal utility of additional resources) it is. The three aspects are defined by the following questions [12]:

- *Importance*: What is the scale of problems in the area? If all problems in the area could be solved, how much better would the world be?
- *Tractability*: How solvable are the problems in this area?
- *Neglectedness*: How much is already being done in this area?

These questions can then be assessed, either, qualitatively or quantitatively, which allows for a very rough estimation of the expected value of directing additional resources into a focus area. Given the recognition of marginal utility [26] assessments of focus areas are highly context specific and may change over time. At the moment, mainstream thinking in EA generally recognizes three major focus areas on which the movement as a whole is focusing, namely, *global poverty and health*, *animal welfare*, and *improving the long-term future* [27]. However, personal fit and obligations may rightfully push individual EAs or subgroups within EA to engage in and explore other focus areas [28].

On the level of specific interventions, practical ethical decision-making is generally built around cost-effectiveness or cost-benefit estimates. When faced with resource constraints investing resources cost-effectively is simply the most effective thing to do. A challenging topic in cost-effectiveness analyses is the inherent uncertainty but

overwhelming importance of long-term consequences [18, 19]. Within EA there is currently no consensus on how to solve this challenge. Thus, personal judgements become an important factor in choosing the most promising interventions [28].

3 What We Can Learn from Effective Altruism

Although we have only had space to present some of the basic ideas and considerations of EA, we hope to have convinced the reader that it provides a thought-provoking lens onto the world when compared to mainstream thinking in IS. This is not surprising, whereas IS research emerged from pragmatic concerns with how to understand and deal with people interacting with IT in business contexts [2], EA is specifically focused on how to have a positive impact in the world [7]. Consequently, if IS research is striving to increase its positive impact on the world [5], it seems only prudent to reflect upon and adopt from the body of knowledge which is already available in EA. For this paper, we have chosen to discuss two major ideas from EA, which seem especially relevant for IS research to consider, namely, *shared goals, principles and measures*, and *cause neutrality and focus areas*.

3.1 Shared Goals, Principles and Measures

EA movement growth is grounded in an ongoing concern with and support of community development, which includes, but is not limited to, the development and institutionalization of *shared goals, principles and measures* [7, 8, 11]. For example, as Walsham [6] wished for IS, EA has been built around a strongly shared ethical agenda focused on a single core goal [7, 27]: *do the most good with the resources that are available*. Moreover, a set of shared principles (i.e., *engagement for others, scientific mindset, openness, integrity, and collaborative spirit*) is prominently being held up by major institutions in the movement [8] and openly celebrated at major EA conferences. Finally, united by the common goal of finding ways to do the most good and guided by the shared principles, EAs find it in their own best interest to develop, establish, and update shared measures to be able to better understand, benchmark and compare how they can do the most good [11, 29].

We posit that these three facets are major parts of a shared identity of EA that enables a diverse set of people and institutions to critically but constructively engage, coordinate, and cooperate in a wide range of focus areas. In our opinion, EA is a living example of how shared goals, principles and measures can create a movement-wide bond and trust that fosters cooperation even among only loosely connected individuals. For example, it is not uncommon for EAs to personally work on one focus area where they have expertise (or simply earn-to-give) and support work on other focus area financially because they share the same goal and can assess and value the work done by other EAs (e.g., due to shared measures and benchmarks). Furthermore, in such an environment, multiple perspectives, theories, and research approaches can all coexist and thrive if they can demonstrate in how far they address concerns which have not yet been captured or solved by existing work because everyone is working towards the

same goal. For instance, population ethics is a highly-contested field in which several theories co-exist but none seem completely satisfactory [30]. This has given rise to several approaches that try to deal with this challenge, including the novel extension of ethical decision-making to the conditions of moral uncertainty [12]. There is simply no gain in rejecting novel or unusual approaches on nebulous grounds if they demonstrate utility for reaching one's own goals.

Translated to IS research this provides evidence for the view held by Walsham [6] that a strong ethical agenda could provide the foundation for a more cooperative, and impact-driven field. Moreover, EA presents us with an example of important aspects that such a field should clarify and provide (i.e., shared goals, principles and measures). This insight complements recent work on a systematic high impact research model (SHIR) [31] which elaborated on how to build high impact IS research programs. Whereas SHIR focuses on how to conduct a specific research program, we highlight the importance of shared goals, principles and measures for coordinating and prioritizing between possibly diverse research programs. For example, the article on SHIR presents three high-impact research programs but does not attempt to quantify or compare impact. This is not surprising because it is very challenging to compare the impact of research programs if there are no readily shared measures between them. To overcome this challenge, IS researchers could be advised to link their work to the UN Sustainable Development Goals (SDG) [32] or human well-being [29] as broad frameworks for assessing ethically-valuable impact. This could be a first step towards more broadly shared goals, principles and measures in IS research.

3.2 Cause Neutrality and Focus Area Selection

The idea of *cause neutrality* and a focus on systematic, evidence-based focus area selection is another corner stone of the EA movement which IS research may adopt and benefit from. Cause neutrality describes the stance that no cause (i.e., focus area) is by any rule, definition or dogma more important than any other cause, only the consequences of working on a cause should determine our actions [8, 12, 14, 25]. The benefits of cause neutrality include that it enables the most effective use of resources [25] as well as that it creates an expectation of constant adaptation and reprioritization as new evidence is assimilated and views are updated. We argue that this facilitates more effective coordination between individual interests as cause neutrality presents a rational avenue to discuss allocation conflicts and mediate between them.

On the practical side of cause neutrality, EA has generally taken to the ITN framework (see section 2.2) to operationalize cause neutral focus area selection. Similar to how it is used in EA, the ITN framework could be used to guide IS research to focus areas where it will likely have the biggest impact. More interestingly, even results from its current application in EA [27, 33] could already be useful to IS as focus areas with high-expected value for EAs are also likely to be focus areas where IS could have an impact. For example, the EA focus area of *global health and development* provides ample opportunities for impactful IS research (e.g., [34]) that can potentially improve the lives of millions at comparatively low costs. The same goes for the other EA focus areas, *animal welfare* (e.g., IS research could aim to reduce the amount of suffering

caused by factory farming), and *improvement of the long-term future* (e.g., IS research could aim to improve institutional decision making and coordination with novel IT-artifacts).

Altogether, insights from EA suggest that reconsidering focus areas in IS could substantially increase the impact of research. Although business organizations have been the starting point for IS research [1, 2], it need not be its destiny. While several researchers and initiatives have made moves towards expanding the scope and mission of IS research [35–37], there is so far no clear conceptual foundation for comparing and choosing between specific directions and focus areas in IS research. Thus, specific initiatives may largely be directed by ambient factors such as social capital of the researchers involved or current trends in other fields and less in logical reason and expected value in any ethical sense. As discussed, the ITN framework can usefully inform future IS research on how to maximize impact by focusing on the most pressing challenges.

4 Discussion and Outlook

In this paper, we have considered EA as a reference discipline that demonstrates the feasibility and utility of a strong ethical agenda for IS research [6]. We contribute to the discourse around IS impact and identity by identifying key considerations that impact-driven IS research should reflect upon and highlight were the existing body of knowledge in EA may usefully inform such research. Importantly, applying EA concepts and ideas could help IS research convert its latent comparative advantage in investigating and shaping IT-related phenomena into often missed “real world” impact [6, 31] by focusing on the most ethically-valuable opportunities. Examples for potentially highly impactful IS topics from an EA perspective are: governance for the safe development of AGI or IS to improve institutional decision-making.

However, our work is not without limitations. While we have started to engage with the considerations we identified, future research is needed to further investigate possible implications for IS research and flesh out more concrete strategies of action. Moreover, due to space limitations, we could only touch on but not fully unpack what EA can offer to IS research. The interested reader is invited to investigate the cited sources to get an in-depth understanding of the topic. Future research should provide more in-depth examples and discussions to make the topic more approachable.

Considering these possibilities and limitations, we encourage other researchers to work with us towards realizing IS research’s full potential to do good in the world. Next steps could include, but are not limited to:

- more open discussions of core goals and values of IS research (e.g., does IS research have a moral *obligation* to strive for the realization of its full potential to do good?),
- the investigation and adaptation of shared measures to IS research (e.g., investigate if the SDG [32] or well-being [29] can be used as frameworks to compare different focus areas, for instance, “health IT” vs. “fintech” vs. “green IT”), or
- a comprehensive application of the ITN framework to focus areas of IS research (e.g., are there focus areas with much higher expected ethical-value than others?).

References

1. Keen, P.G.: Relevance and rigor in information systems research: improving quality, confidence, cohesion and impact. In: Nissen, H.-E., Klein, H.K., and Hirschheim, R. (eds.) *Information systems research: Contemporary approaches and emergent traditions*. p. 49 (1991).
2. Bhattacherjee, A.: Understanding and Evaluating Relevance in IS Research. *Communications of the Association for Information Systems*. 6, (2001).
3. Benbasat, I., Zmud, R.W.: The identity crisis within the IS discipline: Defining and communicating the discipline's core properties. *MIS quarterly*. 183–194 (2003).
4. Teo, T.S.H., Srivastava, S.C.: Information Systems (IS) Discipline Identity: A Review and Framework. *Communications of the Association for Information Systems*. 20, 29 (2007).
5. Wiener, M., Saunders, C., Chatterjee, S., Dennis, A., Gregor, S., Mähring, M., Mertens, P.: Information Systems Research: Making an Impact in a Publish-or-Perish World. *Communications of the Association for Information Systems*. 43, (2018).
6. Walsham, G.: Are we making a better world with ICTs? Reflections on a future agenda for the IS field. *J Inf Technol*. 27, 87–93 (2012).
7. MacAskill, W.: *Doing good better: Effective altruism and a radical new way to make a difference*. Guardian Faber Publishing (2015).
8. Centre for Effective Altruism: CEA's Guiding Principles, <https://www.centreforeffectivealtruism.org/ceas-guiding-principles/>.
9. Future of Humanity Institute: Future of Humanity Institute, <http://www.fhi.ox.ac.uk/>.
10. Global Priorities Institute: Global Priorities Institute, <https://globalprioritiesinstitute.org/>.
11. GiveWell: GiveWell | Charity Reviews and Research, <https://www.givewell.org/>.
12. Centre for Effective Altruism: Effective Altruism Concepts, <https://concepts.effectivealtruism.com/>.
13. Singer, P.: Famine, affluence, and morality. *Philosophy & public affairs*. 229–243 (1972).
14. Jollimore, T.: Impartiality. In: Zalta, E.N. (ed.) *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University (2018).
15. Karnofsky, H.: Excited altruism, <https://blog.givewell.org/2013/08/20/excited-altruism/>.
16. Bostrom, N.: Are we living in a computer simulation? *The Philosophical Quarterly*. 53, 243–255 (2003).
17. Bostrom, N.: The future of humanity. In: *New waves in philosophy of technology*. pp. 186–215. Springer (2009).
18. Beckstead, N.: ON THE OVERWHELMING IMPORTANCE OF SHAPING THE FAR FUTURE, (2013).
19. Todd, B.: How Important are Future Generations?, <https://80000hours.org/2013/08/how-important-are-future-generations/>.
20. Bostrom, N.: Superintelligence. Dunod (2017).
21. Bostrom, N.: Existential Risk Prevention as Global Priority: Existential Risk Prevention as Global Priority. *Global Policy*. 4, 15–31 (2013).
22. J-PAL Policy Bulletin: Roll Call: Getting Children into School. Abdul Latif Jameel Poverty Action Lab, Cambridge, MA, USA (2017).
23. Ord, T.: The Moral Imperative toward Cost-Effectiveness in Global Health. (2013).

24. Tomasek, B.: Why Charities Usually Don't Differ Astronomically in Expected Cost-Effectiveness, <https://reducing-suffering.org/why-charities-dont-differ-astronomically-in-cost-effectiveness/>.
25. Baumann, T.: The Benefits of Cause-Neutrality, <http://s-risks.org/the-benefits-of-cause-neutrality/>, (2018).
26. The Library of Economics and Liberty: Margins and Thinking at the Margin, <https://www.econlib.org/library/Topics/College/margins.html>.
27. Centre for Effective Altruism: Introduction to Effective Altruism, <https://www.effectivealtruism.org/articles/introduction-to-effective-altruism/>.
28. Global Priorities Project: How to help the world, <http://globalprioritiesproject.org/wp-content/uploads/2015/09/Flowchart-medium.jpg>.
29. Plant, M.: A Happiness Manifesto: Why and How Effective Altruism Should Rethink its Approach to Maximising Human Welfare, <https://forum.effectivealtruism.org/posts/FvbTKrEQWXwN5A6Tb/a-happiness-manifesto-why-and-how-effective-altruism-should>.
30. Arrhenius, G., Ryberg, J., Tännsjö, T.: The Repugnant Conclusion. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University (2017).
31. Nunamaker, J.F., Twyman, N.W., Giboney, J.S., Briggs, R.O.: Creating High-Value Real-World Impact Through Systematic Programs of Research. MIS Quarterly. 41, 335–351 (2017).
32. Griggs, D., Stafford-Smith, M., Gaffney, O., Rockström, J., Öhman, M.C., Shyamsundar, P., Steffen, W., Glaser, G., Kanie, N., Noble, I.: Sustainable development goals for people and planet. Nature. 495, 305 (2013).
33. Wiblin, R.: How to compare different global problems in terms of impact?, <https://80000hours.org/articles/problem-framework/>.
34. Ginige, T., De Silva, L., Walisadeera, A., Ginige, A.: Extending DSR with Sub Cycles to Develop a Digital Knowledge Ecosystem for Coordinating Agriculture Domain in Developing Countries. In: Chatterjee, S., Dutta, K., and Sundarraj, R.P. (eds.) Designing for a Digital and Globalized World. pp. 268–282. Springer International Publishing (2018).
35. Lee, J.K.: Research framework for AIS grand vision of the bright ICT initiative. MIS Quarterly. 39, (2015).
36. Lee, J.K.: Invited commentary—Reflections on ICT-enabled Bright Society research. Information Systems Research. 27, 1–5 (2016).
37. Brocke, J. vom, Watson, R.T., Dwyer, C., Elliot, S., Melville, N.: Green information systems: Directives for the IS discipline. Communications of the Association for Information Systems. 33, (2012).