Effective Altruism

How Can We Best Help Others?

Ratio Ethica

Dedicated to David Pearce

Table of Contents

Preface

Two meanings of the term "effective altruism" are worth distinguishing: effective altruism as the ideal of helping others as much as possible, and effective altruism as a social movement consisting of people who are trying to live up to this ideal. This book is about effective altruism in the first sense; it seeks to explore how we can best help others. The book does not speak for the entire effective altruism community, nor could it. Indeed, no one person can speak for a community with such diverse views as the effective altruism community. Yet that should not prevent any member of this community, or indeed anyone else, from publicly exploring what the ideal of effective altruism entails. On the contrary, such critical examination seems necessary if we are to advance toward a more well-considered understanding of the ideal of effective altruism.

As an examination of this ideal, this book should provide food for thought both to people who have never heard the term "effective altruism", as well as to those who think about effective altruism every day.

Introduction: What Is Effective Altruism?

"Effective altruism is the project of using evidence and reason to figure out how to benefit others as much as possible, and taking action on that basis."

— William MacAskill[1]

Effective altruism is all about taking maximal advantage of our enormous opportunity to improve the world. It is about using our limited resources to help sentient beings as much as we can.

There are two core elements to effective altruism. First, there is the altruism itself. This is not altruism in the irresponsible sense of sacrificing for others without any regard for oneself, but rather in the simple sense of improving the lives of others — something that requires us to take good care of ourselves indeed. This may be considered the emotional part of effective altruism, the burning flame at the center that animates the project. Second, there is the effectiveness. This is the cooler, more cerebral aspect that asks us to optimize our core goal: helping other sentient beings is great, but helping them as much as our resources allow is even greater.

In this way, effective altruism is a project that requires us to combine both the heart and the head; empathy and evidence; compassion and careful consideration.

0.1 Singer's Shallow Pond

A thought experiment often used to introduce the motivation behind effective altruism is Peter Singer's shallow pond argument. Imagine that you are walking past a shallow pond in which a child is drowning, and you can save the child if you are willing to jump in the pond and have your clothes and shoes ruined. The question that Singer asks us is: should you jump in the pond and save the child?

This can almost seem like a trick question, as the answer seems an all too obvious "yes". But what if there are other people around the pond besides you who could also save the child, yet who choose not to? That would not seem to change much; most of us agree we should still save the child. Imagine, then, that the drowning child is not in a pond right next to you, but a full

---

1 MacAskill, 2017, p. 2.

kilometer away, and imagine that you can still save the child by exercising the same amount of effort that would be required if the child were in a pond right next to you. Should you then still save the child? The answer, again, seems an obvious "yes". So too if the child were 100 kilometers away, or indeed on the other side of the planet, provided that the child can still be saved with the same low level of effort.

All of this can seem rather unremarkable: we have merely established the rather obvious proposition that we should save a child's life, regardless of where in the world the child is, if we can do so at a minimal cost to ourselves, such as the price of some clothes and a pair of shoes. Yet the point of Singer's argument is that the implications of accepting this proposition are in fact anything but trivial, since we, Singer argues, find ourselves in a similar situation right now:

> We are all in that situation of the person passing the shallow pond: we can all save lives of people, both children and adults, who would otherwise die, and we can do so at a very small cost to us: the cost of a new CD, a shirt or a night out at a restaurant or concert, can mean the difference between life and death to more than one person somewhere in the world — and overseas aid agencies like Oxfam overcome the problem of acting at a distance.[2]

So how can we defend not donating to charities that do such life-saving work, given that we agree that saving someone's life is worth the price of some clothes and a pair of shoes? One may, of course, object that there is greater uncertainty in the case of organized charity, yet as Singer notes:

> Even if a substantial proportion of our donations were wasted, the cost to us of making the donation is so small, compared to the benefits that it provides when it, or some of it, does get through to those who need our help, that we would still be saving lives at a small cost to ourselves — even if aid organizations were much less efficient than they actually are.[3]

2 Singer, 1997.
3 Singer, 1997.

It should be noted that the specifics of this thought experiment have been criticized by many, including some self-identified effective altruists. For instance, the price of saving a human life does not seem comparable to that of a CD, but appears to instead be in the thousands of dollars.[4]

Yet irrespective of any particular criticism we may level at Singer's argument, the more general point still stands undisputed. We, as citizens of the world, have the potential to greatly help other individuals, likely a large number of them, with the time and money that we have at our disposal, and many of us can do so without sacrificing anything of comparable value to ourselves. The core question of effective altruism is how we can best realize this potential. That is the question this book seeks to examine.

0.2 A Brief Note on Ethics

A preliminary question worth clarifying is whether effective altruism is just a rebranding of utilitarianism, the ethical theory that requires us to bring about the greatest wellbeing — or the least illbeing — for sentient beings. The short answer to this question is "no". For although utilitarianism implies that we should be effective altruists, at least of some kind, the arrow does not point the other way. That is, one can be an effective altruist without being a utilitarian.

For instance, one can be an ethical pluralist who ascribes value to a wide variety of things, where helping others effectively with one's surplus resources is one of them. Indeed, one can be a deontologist or virtue ethicist and consider effective altruism a natural consequence of these views.[5] In particular, one may consider it a duty to follow the rule "try to help others effectively with your surplus resources", or consider efforts to help others in this way an essential part of a virtuous life.

---

4 Sinick, 2013.

5 Deontology, or duty ethics, refers to a class of ethical theories that hold that we should always act according to certain duties or rules — e.g. Immanuel Kant was a deontologist who thought we should always act according to the rule "do not lie". Virtue ethics, which is mostly associated with Aristotle, holds that we should try to be as virtuous as possible, e.g. embody virtues such as modesty, honesty, discipline, and compassion as well as we can in our daily lives.

1. The Core Virtues of Effective Altruism

Before we ask how we can maximize our potential to help others, let us first review some ideas and thinking tools that play a crucial role in the pursuit of effective altruism — one could even say that they are its defining virtues.

1.1 It Is an Open Question

The first and most foundational of these virtues is to admit that the question concerning how we can best help others really is a question, and an extremely open and complex one at that. This may seem obvious, yet our attempts to improve the world nonetheless rarely reflect this openness and complexity. Most people have their own ideas about how to best improve the world, but few seem to have reflected critically upon such proposals. This should not come as a surprise, as we did not evolve to ask and think deeply about questions of this nature, much less come up with plausible answers to them. On the contrary, we arguably evolved a strong drive to mostly affirm the answers of our peers, and to proclaim answers that make us *look* good.[6] Yet such natural inclinations only highlight the need to engage in critical examination and to resist the ever-present temptation of accepting easy answers.

1.2 Impartiality

Impartiality refers to the principle that we should prioritize equal interests equally. In other words, the principle of impartiality entails that it is the sentience of an individual — an individual's capacity to experience states of happiness and suffering — that makes that individual worthy of our moral concern and help, while other criteria, including an individual's gender, sexual orientation, species, and position in space and time, are not relevant per se.

1.3 Dedication to Reason

Being dedicated to reason means being willing to follow arguments and evidence, wherever these may lead us. One can argue that the principle of impartiality described in the previous

---

6 Cf. Simler & Hanson, 2018, ch. 12.

section follows directly from such a dedication, because whatever our conception of reason may be, consistency must at the very least be considered an integral part of it. And it is not consistent to treat the same thing differently — the same amount of suffering, say — depending on where in time and space it happens to be instantiated. That would be like saying that 2 + 2 is 4 in my head, yet not necessarily in the head of any other person. As utilitarian philosopher Henry Sidgwick wrote in his *Methods of Ethics*, "Reason shows me that if my happiness is desirable and a good, the equal happiness of any other person must be equally desirable." (Book 3, Chapter 14.)

Another core virtue that follows from a dedication to reason is that of cause neutrality. Cause neutrality is simply the rejection of favoritism at the level of the altruistic cause we dedicate our resources to. We should not favor, say, the alleviation of human poverty over other causes as being most important just because we happen to feel a strong attachment to it, or because we have put a lot of effort and resources into it in the past. Instead, both when it comes to different cause areas, as well as to specific interventions within these cause areas, we should remain open-minded and impartial so as to be able to follow the evidence as best we can.

1.4 Being Aware of Biases

Research over the last few decades has revealed that our thinking is systematically biased in various ways.[7] Therefore, if our views and actions are to be based on reason and evidence, we must be careful to avoid these pitfalls of the human mind, the stumbling blocks to reason. I shall not review the literature on such biases here, but instead just mention a few of the most important biases that we should be aware of, and which I will refer back to later in this book.

1.4.1 Confirmation Bias

Confirmation bias is the well-documented tendency of our minds to search for, interpret, and recall information that confirms our preexisting beliefs and opinions.[8] This is obviously not a recipe for forming reasonable beliefs. The way to counteract this prevalent bias, it would seem, is to become acutely aware of it in the first place. We can then try to counter it by suspending our attachment to our established beliefs, and by seeking out and giving fair hearing to arguments and viewpoints that may contradict our favored ones. In other words, we can try to keep an open

---

7 A popular book on the subject is Daniel Kahneman's *Thinking, Fast and Slow.*
8 Plous, 1993, p. 233.

mind. This is, of course, much easier said than done. Yet it is crucial just the same — for forming good beliefs in general, and for the project of effective altruism in particular.

1.4.2 Overconfidence Bias

A bias closely related to confirmation bias is overconfidence bias: our tendency to have excessive confidence in our own beliefs. For example, in one study, people who reported to be 100 percent certain about their answer to a question were only correct 70 to 85 percent of the time.[9] This bias should make us pause with respect to claims we profess to have 100 percent certainty in, and indeed with respect to any confidence we may have (had) in the reliability of the human mind in the first place. In the context of the pursuit of effective altruism, it should make us question even the most confident of claims we believe about which actions will help others the most. The remedy for this bias, it seems, is to compensate by weaving some humility into our credences, and by embracing intellectual humility more generally.

1.4.3 Wishful Thinking

Another tendency closely related to confirmation bias is our tendency to engage in wishful thinking.[10] That is, our tendency to believe what we wish were true. Recognition of this tendency should make us skeptical about those of our beliefs that happen to match up with what we wish were true — for instance, that the future will be wonderful, that life for most beings on Earth is likely to be good, or that the worst forms of suffering cannot be *that* bad.

Likewise, our tendency to engage in wishful thinking suggests that it is helpful to pay more attention to ideas that are inconvenient or otherwise contrary to what we wish were true, such as the notion that a behavior we are engaging in is morally indefensible, or that the world contains far more horror than we had ever imagined.

1.4.4 Groupthink

Groupthink refers to a phenomenon where a lack of dissenting viewpoints and critical perspectives within a group leads to irrational beliefs and decisions among its members.[11] It is the product of our general tendency to conform to our peers, not only in how we behave, but also in

---

9 Plous, 1993, pp. 219-220.
10 Bastardi et al., 2011.
11 Plous, 1993, pp. 203-204.

what we believe.[12] This conformist tendency should make us skeptical of the beliefs that we happen to share with our own "group" — our close peers as well as our culture at large — and it gives us reason to practice playing the devil's advocate against those beliefs.

1.4.5 Scope Neglect

This bias is also called scope insensitivity: our feelings about a disaster and our willingness to donate toward its alleviation appear largely insensitive to the number of individuals it affects.[13] This is related to the identifiable victim effect: we generally offer greater help when we are presented with an identifiable victim of a disaster than when we are presented with facts about how many individuals it affects. And as psychologist Paul Slovic has shown, not only does our empathy not increase as the magnitude of a disaster grows, but it actually seems to decrease rapidly.[14]

This is obviously not reasonable: a disaster that involves more beings is, other things equal, worse than one that involves fewer beings. Yet our moral sentiments somehow turn this relationship completely upside-down, which again has important implications for any aspiring effective altruist. Most significantly, it implies that we should be skeptical of our immediate feelings about which disasters and risks we should prioritize the most, and it suggests that we should make an effort to consider the number of individuals involved. This may seem cold, as though we decide by considering numbers rather than individuals. Yet it is actually the opposite. For the numbers must be considered exactly because they represent individuals — individuals who would otherwise be ignored by our faulty moral intuitions. Therefore, it is to *not* consider the numbers that ultimately amounts to a failure to consider the individuals. As Michael Moor put it in his excellent TED talk "Impact through rationality", numbers count *because* individuals count.[15]

1.4.6 Speciesist Bias

The term "speciesism" refers to unjustified discrimination against beings based on their species membership.[16] And if we are committed to the principle of impartiality mentioned above, we are committed to rejecting *all* forms of unjustified discrimination. In particular, we will be

---

12 Plous, 1993, pp. 203-204.
13 Animal Ethics, 2017.
14 Slovic, 2007.
15 Moor, 2013.
16 Horta, 2010; Vinding, 2015.

committed to the position that, as Peter Singer puts it in the *Effective Altruism Handbook*, "'speciesism,' by analogy with racism, must also be condemned."[17]

Not surprisingly, given an evolutionary history in which killing beings of other species was necessary for human survival, and given that we now find ourselves in a modern society in which we have completely trivialized the exploitation and killing of non-human beings, the anti-speciesist position is one that we have a hard time accepting. In other words, we humans have an enormous speciesist bias. As a study led by Lucius Caviola at the University of Oxford reported: "Our studies show that people morally value individuals of certain species less than others even when beliefs about intelligence and sentience are accounted for."[18]

Another study found that we are inclined to ascribe diminished mental capacities to, and even entirely deny the existence of a mind in, the non-human individuals whom we consider food. And this mind denial is increased by "expectations regarding the immediate consumption" of such beings.[19] In other words, eating a group of beings appears to make us biased with respect to the capacities we ascribe to those beings, which in turn biases our moral cognition about them.

1.5 Expected Value Thinking

The final virtue of effective altruism I shall review here is expected value thinking. Such thinking reflects the insufficiency of evaluating which possible outcomes to aim for merely in terms of the probability *or* the value of these outcomes. We have to factor in both: the estimated probability of bringing about a given outcome *multiplied by* its estimated value. This is the expected value of trying to realize this given outcome.

To take a concrete example involving two betting options: say for option A we have a 50 percent probability of winning a prize of 10,000 dollars, which means that option A has an expected value of 0.5 times 10,000 dollars (= 5,000 dollars), while option B gives us a one percent chance of winning a prize of a million dollars, which yields an expected value of 0.01 times 1,000,000 dollars (= 10,000 dollars). So although the probability of winning the prize is much smaller with option B, we should still bet on that over option A on the expected value framework (assuming that our aim is to win the greatest amount of money).

---

17 Carey, 2015, p. 98.
18 Caviola et al., 2019, abstract.
19 Bastian et al., 2012.

This example illustrates that even unlikely outcomes can still be the most important ones to focus on, provided that they are sufficiently valuable. And conversely, even very valuable outcomes may not be the most promising ones to focus on, if they are sufficiently unlikely.[20]

A specific framework often employed by effective altruists that in some sense concretizes the expected value framework is the scale, tractability, and neglectedness framework, which can help us evaluate the altruistic potential found within a given problem.

Scale, or scope, refers to the size of the problem. There is no unique measure for evaluating scale, but it could, for example, be in terms of how many individuals it affects, or how intense it is (e.g. intense suffering can reasonably be said to have great magnitude even if it only affects relatively few individuals). **Tractability**, sometimes called solvability, refers to how amenable to solution the problem is. That is, how much of the problem do we expect to solve per resource we spend on it? Finally, **neglectedness** is how overlooked the problem is by other people. This last consideration is relevant because a neglected problem will be more likely to have low-hanging fruits ready to be picked compared to a problem that many people are working on. Or expressed in more general terms, our efforts on more neglected problems are less likely to be subject to (as strong) diminishing returns compared to efforts on less neglected problems. Considering these three factors in combination may be helpful to get a rough sense of how much we are able to help others by spending our resources on a given problem.

Closely related to the concept of maximizing expected value is the concept of opportunity cost: the loss of potential gain from other alternatives when one alternative is chosen.[21] To choose one opportunity is to forego another, and in trying to maximize the expected value of our actions, we are essentially trying to minimize our opportunity cost by choosing the best opportunity that is available to us. Thinking in terms of opportunity costs forces us to think carefully about the alternatives to our plans, and to consider whether those alternatives might be better.

It should be noted that expected value calculations are rarely exhaustive in a good decision-making process, whether altruistic or not. We cannot always quantify the value or the probability of a given outcome with great precision, and hence other styles of thinking should often weigh stronger in our decision-making.[22] Still, the expected value framework can have great utility, and is worth including in our considerations and decision-making when possible.

---

20 For a case for why we should maximize expected value, see Tomasik, 2007.
21 As defined by the *Oxford American Dictionary*. In economics it is defined more specifically as the value of the best option not chosen.
22 See also Karnofsky, 2011; 2014.

With these principles and thinking tools in mind, let us proceed in our examination of the core question of effective altruism: how can we best help others?

2. Monetary Donation

The act of giving charitable donations is perhaps the first thing that springs to mind when it comes to actions we can take to help others in need. And it turns out that such donations can go a surprisingly long way. Let us review some estimates of just how far they can go in different spheres.

2.1 Human Poverty

The organization GiveWell evaluates charities that help humans in the developing world in order to find out which of these charities achieve the most per dollar donated. GiveWell cites two reasons for this focus on people in the developing world: 1) poor people in richer countries are "wealthy by developing-world standards", and 2) probably for this first reason, GiveWell has found no poverty-targeting interventions in richer countries that compare favorably to its most recommended interventions in the developing world.[23]

One of the charities that GiveWell recommends most highly is the Against Malaria Foundation (AMF), a charity that enables the distribution of cheap bed nets that have been shown to reduce child mortality and malaria cases in developing countries. On GiveWell's estimates, a human death from malaria is prevented for roughly every 5,500 US dollars donated to AMF.[24] Other charities recommended as highly effective by GiveWell at the time of writing include the Schistosomiasis Control Initiative, which treats people for painful parasite infections in sub-Saharan Africa; Evidence Action's Deworm the World Initiative, also treating parasitic infections; and the Helen Keller International vitamin A supplementation program, doing what the name suggests, i.e. providing vitamin A supplementation, which is a cheap way to reduce child mortality.

In a world where more than five million children under the age of five die every year, and where more than half of these deaths "are due to conditions that could be prevented or treated with access to simple, affordable interventions", it is difficult to defend not doing *something*.[25] And upon deciding to do something, it seems difficult to defend not examining how we can do the

23 GiveWell, 2016/2021.
24 GiveWell, 2012/2022; 2021/2022.
25 WHO, 2020.

best something. It would be tragic to fail to save a life by spending one's altruistically devoted resources in a suboptimal way. The imperative of avoiding such tragically suboptimal actions is, of course, not unique to the cause of human poverty, but applies to the endeavor of effective altruism in general.

2.2 Meta Charity

Meta charities include organizations that evaluate charities, such as GiveWell mentioned above, as well as the organization Giving What We Can (GWWC), which also focuses on helping people in the developing world. By donating to such charities rather than to the charities they recommend, one should expect to improve the quality of their recommendations. Whether this will be the optimal decision depends on how much better one expects these evaluators' recommendations to be as a result of the added resources, which should then be compared to what the evaluated charities could do with the resources. Yet it also depends on whether the evaluator and evaluated charities are focusing on the most promising cause in the first place, e.g. global poverty as opposed to, say, helping non-human animals.

An organization that does charity evaluation within the latter cause — advocating for the interests of non-human individuals and sparing them from unnecessary suffering and death — is Animal Charity Evaluators (ACE), whose recommendations we shall take a closer look at now.

2.3 Helping Non-Human Beings

The vast majority of sentient beings on the planet, more than 99.99 percent of them, are non-human.[26] In spite of this, virtually none of humanity's charitable donations are devoted to this group of beings. These two facts alone strongly suggest that we can help non-human individuals a lot with our limited resources, a suggestion that is borne out by more elaborate estimates.

A 2018 estimate from ACE suggested that one can expect a donation of 1,000 US dollars to the best charities within this cause to spare more than 4,000 beings from a life on a factory farm.[27] This is an enormous number, so large that we, in order to get a sense of its moral magnitude, probably need to remind ourselves of our scope neglect — our inability to appreciate the difference between helping 1, 10, 100, 1,000, and 4,000 individuals. More than that, we should

---

26 Tomasik, 2009c.
27 Animal Charity Evaluators, 2018.

also expect our moral intuitions to fail to appreciate the value of helping so many such beings due to our speciesist bias.

Some of the charities that ACE recommends are The Humane League, Anima International, and Mercy for Animals. Another charity that deserves mention, and which ACE has recommended as a standout charity in the past, is Animal Ethics. This charity is unique in that it focuses on advocating for *all* non-human beings, including those living in nature, which turn out to be the vast majority of sentient beings on the planet.[28]

2.4 Effective Altruism Research

Another potentially promising donation opportunity is that of funding research into effective altruism — that is, research on how we can best help others. The work done by the charity evaluators mentioned above is an example of such research. Yet evaluating the near-term effects of existing charities is by no means the only form of research that can help inform our priorities.

Some research centers have a stronger focus on how we may influence the long-term future in beneficial directions. These include the Global Priorities Institute (GPI), the Center on Long-Term Risk (CLR), and the Center for Reducing Suffering (CRS). (I should note that I am a co-founder of CRS, together with Tobias Baumann.)

The Global Priorities Institute conducts broad foundational research on altruistic prioritization, mostly from a perspective that is quite optimistic about humanity's future. In contrast, the Center on Long-Term Risk focuses more narrowly on reducing worst-case risks due to advanced AI systems, whereas the Center for Reducing Suffering pursues a broad research project on how we can best reduce suffering in the long-term future.[29]

It is difficult to put numbers on the value of such research into effective altruism. There is no easy way to make a controlled experiment that assesses whether work of the kind pursued by GPI, CLR, and CRS is cost-effective or not. The long-term-focused nature of such research inevitably makes evaluations of its value more difficult, compared to, say, zooming in on an intervention and asking what its immediate effects are. Yet this large-scale focus is also what makes such research promising. Not many people are doing such work, and if crucial considerations that force our altruistic priorities to change altogether emerge as the result of it, then it could well turn out to be exceptionally important.

---

28 Tomasik, 2009c.
29 See globalprioritiesinstitute.org; longtermrisk.org; centerforreducingsuffering.org

We will probably never be able to answer with great confidence which donation opportunities are best with respect to the goal of helping sentient individuals as much as possible. Yet to the extent that we can answer it with at least some confidence, such an answer is going to depend upon deeper considerations than any we have raised so far, including considerations regarding what helping others means in more precise terms. These are considerations that we shall delve into in later chapters.

3. Career Choice

A significant part of our waking lives will be spent doing professional work, which renders our choice of career a critically important one with respect to our potential to help others.

The organization 80,000 Hours, whose name derives from the number of hours the average person can expect to spend on their career, was founded with the purpose of helping people find a career with great altruistic potential. What such a career might be for any given individual is, of course, a difficult question, and again one we will likely never have definitive answers to. Yet there still seem to be some informed principles that are worth following in our attempts to maximize the altruistic impact of our careers.

One such principle that 80,000 Hours emphasizes is "don't follow your passion". This may seem counterintuitive, and not least contrary to everything found in the modern canon of popular motivational speeches. Yet the point is not that one should not be highly passionate about what one does — that is indeed recommendable, probably even necessary. Rather, the point is that our passions can change over time, and the thing we are most passionate about at the moment might not be what we will be most passionate about in the future, and it might not be something that helps other individuals much, let alone what helps others the most.

So instead of pursuing our immediate passions, 80,000 Hours recommends that we make a careful analysis of which career paths seem to have the best altruistic impact, and then consider pursuing the one that seems most promising — keeping in mind, of course, that different individuals have different talents and dispositions that to a great extent determine what careers they can be most successful in.

A related consideration is how fulfilling one would find a given career path. This is relevant for many reasons: it can be difficult to be altruistically motivated if one is unfulfilled; an unsatisfying career path is less sustainable than a satisfying one (and thus the risk of burnout is higher); and we are generally more productive in our work the happier we are.[30]

Also relevant are the counterfactuals: what would happen if you did not take a given career path? Would someone else take it, and if so, would they have a better or worse altruistic impact through that path compared to if you had pursued it? If they would have a better impact, or even

---

30 Tenney et al., 2016.

an equally good one, it might be better for you to choose another path where you would make a greater difference, all things considered.[31]

So when we make a deeper analysis, we see that the question is not just, "What job that helps others a lot could I do well?", but also, "What job that helps others a lot could I do *particularly* well?" and "What difference can I make that would not have happened otherwise?" In this way, the analysis quickly becomes rather complex, dependent not just upon the countless opportunities and talents that we have ourselves, but also on the opportunities, talents, and choices of other people.

3.1 Earning to Give

An idea often associated with effective altruism is that of pursuing a high-earning career in order to donate a lot of money to charity, commonly referred to as Earning to Give.

At the first level of analysis of the counterfactuals, this appears to be a promising path. An altruistic person who takes a job at a hedge fund in order to donate a large fraction of the salary likely has great value compared to someone else taking that same job, as that someone else likely would not donate nearly as much, if anything. In this way, a person who pursues such a path may be able to fund the salary of several people working at a non-profit organization. Compare this to the counterfactual value of person A choosing to take a job in a non-profit that would otherwise have been taken by person B. The difference between the work that person A and person B could do would have to be large in order for it to be bigger than the difference that could be made by several additional employees funded by a high-earning person doing Earning to Give.

So Earning to Give appears to be a promising path. Yet there are also counter-considerations suggesting that it might not be ideal, at least for many people. One is that of personal fit: a lucrative career may afford a large altruistic impact, yet most people cannot earn an extremely high salary, and even if one could get a high-earning job, working such a job may not be suitable if the level of personal fit is too low (e.g. one might not find it motivating or sustainable in the long term).

Another consideration against Earning to Give is that many non-profit organizations, including most of the major ones that identify as being part of the effective altruism movement, report being significantly more constrained by talent than by funding. In other words, getting more competent people to work for these organization is a greater bottleneck for the progress of their

---

31 And this can also hold true even if they would have a significantly worse impact than you, cf. the concept of comparative advantage.

work than is a lack of money. This is not to say that Earning to Give is not important, or even that it is not neglected by aspiring altruists, but rather that there are other things that appear even more neglected. For example, in one survey, the three most wanted skills at organizations in the effective altruism movement were, in order of importance, 1) good calibration, wide knowledge and ability to work out what is important; 2) generalist research skills; and 3) management skills.[32]

With the competence of being widely read and clear-thinking listed as both number one and two, it seems that undertaking the classical virtues of reading widely and thinking deeply about what is most important, and doing so full-time if possible, might generally be more beneficial than is Earning to Give — especially if one is well-suited for such things. And it can, of course, be reasonably argued that it is worthwhile to practice such virtues in any case.

In sum, our choice of career is a consequential one for our future altruistic potential, and which career is best depends on many things, including our own talents and dispositions, as well as what other people are and could be doing. It also depends on the deeper considerations alluded to at the end of the previous chapter concerning questions of a more fundamental nature, questions that we shall turn to soon.[33]

---

32 Todd, 2015; Wiblin, 2017.
33 For more on altruistic career choice, see 80000hours.org; Baumann, 2022.

4. The Long-Term Future

The core effective altruist principle of impartiality requires us to include *all* sentient beings in our moral considerations, regardless not only of their race, sexuality, and species, but also where in time and space they find themselves. In particular, it implies that we must take the moral value of future individuals seriously, especially since the number of beings who will exist in the future will likely be much greater than the number of beings alive today. This has profound implications for the project of effective altruism, and forces us to consider everything we have seen so far in a new light. It suggests that the main criterion for evaluating how much a given action helps other individuals may be its expected impact on the long-term future.

If we consider the long-term future as a cause to focus on in light of the scope, tractability, and neglectedness framework, we see that it does exceedingly well on at least two of these criteria. Foremost of these is the scope of the cause, which could barely be greater: we should expect the vast majority of sentient beings whom our actions can influence to exist beyond the next couple of decades that usually steal most of our altruistic attention. And this short altruistic attention span then leads us to the neglectedness of the long-term future, which is likewise significant, for various reasons.

First, as creatures who evolved in an environment where it was generally costly to move any attention away from the near future, it is not surprising that we pay less attention to the far future than deeper philosophical reflection might recommend. Second, beyond our evolved nature, we also live in a social world that does not strongly incentivize us to make an effort to positively impact the future beyond the next few decades. Politicians are incentivized to focus mostly on the next couple of elections; companies generally focus on making profits on a year-by-year basis; and non-profits are likewise under pressure to demonstrate immediate tangible results.

Thus, the long-term future scores very high both in terms of scope and neglectedness, and how such a neglect can exist in spite of the vast scope seems explainable, at least in part, by naturally emerged limitations of our minds and social structures. Yet it may also have something to do with the final, as yet unvisited criterion: the tractability of improving the long-term future. For it may just be really difficult to impact the long-term future in a beneficial way, and in the absence of any concrete suggestions about how we might improve the far future, it should perhaps not be surprising that we are not acting on such suggestions. Yet whether said tractability is really that

low remains an open question. Relatively little attention has been devoted to this question, which suggests that more research should be a high priority.

4.1 Impacting the Long-Term Future: Direct or Indirect Influence?

A useful distinction is that between direct and indirect ways of impacting the far future. A direct way could be to work on a specific technology or social institution that one expects to exist and be influential in the long-term future, while indirect ways include more general actions such as advocating on behalf of the individuals who are worst off today, which may then — in a more indirect, ripple-effect manner — lead to a better long-term future. Effects of this latter kind have been referred to as long-run indirect effects, or flow-through effects.[34]

Whether the best way to impact the far future is generally via direct or indirect efforts stands as an open question. An argument in favor of prioritizing indirect efforts is that the values of future individuals are extremely important, and perhaps the best way to positively impact future values is to push for concrete interventions that promote better values today. By this line of reasoning, it could be that supporting the charities mentioned in the previous chapter focused on relatively short-term betterment, such as AMF or ACE, would also be optimal with respect to the long-term future.

One can, of course, reasonably argue that we should be skeptical of such a convenient convergence between optimal short-term and long-term interventions. Yet one can also reasonably question such skepticism. After all, why should what looks optimal from a short-term perspective not be optimal with respect to the long term too? For example, is it unreasonable to think that a person in the year 1800 who wanted to minimize future racism would have the most success by working on the concrete near-term goal of abolishing race-based slavery?

Regardless of what our position may be on such empirical questions, it is important to avoid the non sequitur that the significance of the long-term future implies that *only* the long-term future matters. That simply does not follow. "It's just near-term suffering, so it doesn't matter that much in the great scheme of things" is not a notion we entertain when it comes to our own suffering, and neither should we entertain such notions when it comes to the suffering of our contemporaries.

---

34 This distinction between direct and indirect ways of impacting the future is, of course, far from clear-cut. It is arguably a spectrum, and many projects and interventions can reasonably be identified as both.

4.2 The Future of Artificial Intelligence

Technical interventions aimed at improving the future of artificial intelligence represent a fairly direct effort to improve the long-term future. The reasons behind this focus are presented most elaborately in the book *Superintelligence: Paths, Dangers, Strategies* by philosopher Nick Bostrom. A short version of the argument goes roughly as follows: The main reason that humans rule the Earth is our intelligence, which means that as we create more cognitively capable systems that pursue their pre-programmed goals more intelligently than humans can, there is a risk that such systems will assume control of the world based on these pre-programmed goals. Therefore, the argument goes, we better make sure we get these goals and their implementation right so as to avoid bad outcomes, an effort that has become known as AI safety.

Whether AI safety is indeed the best way to create a better future has been disputed.[35] For example, economist Robin Hanson has argued that it is highly unlikely that the future will depend crucially on how a single, or even a few, powerful software programs are designed.[36] The argument made by Hanson and others is not that AI safety is unimportant, but rather that there are other problems that appear even more urgent (or similarly urgent yet more neglected). In other words, they are arguing that we should carefully consider the opportunity costs of AI safety work, and that we should explore a broad range of causes before committing too strongly to any one of them.

So while impacting the future of artificial intelligence is considered a promising way to positively influence the future by many effective altruists, there is still much room for debate about what kind of work has the *greatest* expected value. After all, even if the future of artificial intelligence is all-important, this does not imply that we should necessarily focus mostly on software design, as opposed to focusing on, say, the larger system of institutions in which the software will be developed and implemented. At the very least, it seems worth avoiding premature answers in either direction on this issue — e.g. "AI safety is clearly the most important cause" or "AI safety is clearly *not* the most important cause". It is important that we keep our minds open and are willing to explore other causes and ways to impact the long-term future when we know so little.

A significant caveat with respect to focusing on the long-term future is that the long-term future tends to be more difficult to reliably influence than the near-term future. This obviously does not

---

35 See e.g. the reading list compiled in Vinding, 2017.
36 Hanson, 2014.

render the long-term future unimportant. It still seems that more research into the question of how to best impact the long-term future should be among our top priorities. Yet we should be realistic about what to expect: highly uncertain answers, at best of a modestly qualified kind.

So while the impact our actions have on the long-term future may well be the most important thing about them, it is far from clear how we can have the best long-term impact. And much more fundamentally, before we can begin to say what is best in the long term — or the short term for that matter — we must have greater clarity about what "best" even means. This is what we shall turn to now.

5. The Cruciality of Clarity About Values

What we have seen so far in our examination of how to best help others has been of a rather general nature, in two senses. First, it is general in that the ideas and considerations we have visited are ones that people in the effective altruism movement tend mostly to agree with. Second, it is general in that we have phrased our project in rather generic terms: how can we best help, or benefit, others? We have said nothing about what benefiting or helping others means in more specific terms, nor about what constitutes the greatest such help. While this was perhaps permissible for the purpose of presenting the preceding considerations, it will not do for any serious project of effective altruism. For how can we systematically try to help others or improve the world as much as possible if we do not have decent clarity about what this means? In order to prioritize and navigate meaningfully and effectively toward the goal of helping others, we must have greater clarity about what this goal entails.

One might say that specifying the meaning of "helping others" beyond our shared common-sense understanding of it amounts to little more than philosophical hair-splitting. After all, don't we know what helping and benefiting others is when we see it? Yet this sentiment, while intuitive and understandable, is misguided. The problem is that when it comes to helping or benefiting others, common sense is a lot less common than one might be inclined to think. For instance, on the subject of population ethics, we find widely held views that are, at least in some sense, diametrically opposed. On the one hand, there are philosophers who believe that we should bring about as much well-being as possible (the greatest amount of happiness minus suffering, roughly speaking), while on the other, there are philosophers who defend a view referred to as the Asymmetry, which roughly holds that adding miserable lives to the world makes the world worse, whereas adding happy lives to the world does not make the world better. Given such a fundamental disagreement, it is unclear what could meaningfully be considered "common sense" when it comes to helping others.

The need for greater clarity is perhaps less obvious when we only focus on the short term, where we can at least readily identify actions that constitute *some* help to other beings by common sense and consensus. But as we extend our focus to the farther future, the need for ethical clarity becomes ever more pronounced. For example, in light of the above-mentioned divergence on population ethics, it is unclear whether we should aim to drastically increase the number of

sentient beings in the universe so as to maximize happiness for the currently unborn, or whether we should not consider this a priority at all. Thus, even when comparing widely endorsed ethical positions, we readily find enormous differences in terms of their practical implications — e.g. a future full of sentient life versus an empty one — which highlights the cruciality of reflecting deeply on our values.

5.1 Clarifying Our Values

So where do we begin if we are to gain clarity about our values? An obvious place to start might be to consider different theories of ethics. Three prominent classes of such theories are virtue ethics (roughly: we should be virtuous), deontological ethics (roughly: we should act according to certain duties), and consequentialist ethics (roughly: we should bring about the best consequences).

Most self-identified effective altruists lean toward consequentialism. Yet even if we were to declare ourselves pure consequentialists, we would still have a long way to go toward specifying our values, as we would then still need to clarify what *kinds* of consequences, or states of the world, that are of greater value than others. We may do this by saying that we are consequentialists of a utilitarian flavor. Yet this is just a small step toward greater specification, as there are also many different kinds of utilitarianism, and they can differ quite radically from each other in what they hold to be valuable, and hence also in terms of their practical implications.

In its classical formulation, utilitarianism holds that we should bring about the greatest sum of happiness minus suffering, commonly referred to as hedonistic utilitarianism. Another version, preference utilitarianism, holds that we should maximize not happiness, but preference satisfaction minus preference frustration. And these two theories can have radically different implications. Beyond that, there are also negative versions of both these kinds of utilitarianism, which give priority to the minimization of suffering and preference frustration, respectively. And the difference in practical implications between negative and non-negative versions of the same theory can likewise be extreme.

Now, it may be tempting to think that if we specify our view down to the level of granularity we have zoomed in on at this point, then we would be close to having a sufficiently specific view of what has value. Yet this is far from the case. For even if we decide to endorse, say, a version of negative hedonistic utilitarianism — i.e. we prioritize reducing conscious experiences of

suffering — we will still be left with many open questions. For example, how do we prioritize different kinds of suffering? Should we give greater priority to intense suffering than to mild forms of suffering, perhaps so much so that it is impossible for any number of mildly bad experiences to be worse than a single very bad one? These are all crucial questions that we must consider in order to be clear about what we are aiming for.

We urgently need to have a deeper conversation on values and about what it means to help others, since everything else depends on these foundational issues — from the activism of today to the values that we should ideally implement in the AIs of tomorrow. In the hope of inspiring such a conversation, I will now try to lay out a rough account of the values that I myself favor.

6. Suffering-Focused Ethics

The values that I find most plausible fall within a broader class of ethical views called suffering-focused ethics, which encompasses all views that give special priority to the alleviation and prevention of suffering. In this chapter, I will review some general arguments and considerations in favor of such views.

It should be noted, however, that not all effective altruists agree with this view of values. Many appear to view the creation of happiness as having the same importance as the reduction of "equal" suffering. I used to hold this view as well, yet I have changed my mind in light of considerations of the kind presented below.[37]

6.1 The Asymmetries

We have already briefly visited one asymmetry that seems to exist, at least in the eyes of many people, between suffering and happiness, namely the Asymmetry in population ethics. The Asymmetry roughly says that we have moral reason to avoid bringing miserable lives into the world, but no moral reason to bring happy lives into the world.

This view has been defended by philosopher Jan Narveson, who has argued that there is value in making people happy, but not in making happy people.[38] Philosopher Christoph Fehige defends a similar view, called antifrustrationism, according to which there is value in making preferrers satisfied, but no value in making satisfied preferrers.[39] Peter Singer, too, has expressed such a view in the past:

> The creation of preferences which we then satisfy gains us nothing. We can think of the creation of the unsatisfied preferences as putting a debit in the moral ledger which satisfying them merely cancels out. … Preference Utilitarians have grounds for seeking to satisfy their wishes, but they cannot say that the universe would have been a worse place if we had never come into existence at all.[40]

---

37 A more elaborate case for suffering-focused ethics is found in Vinding, 2020a, Part I.
38 Narveson, 1973.
39 Fehige, 1998.

In terms of how we choose to prioritize our resources, there does indeed — to many of us at least — seem to be something frivolous about focusing on creating happiness de novo rather than on alleviating and preventing suffering first and foremost. As Adriano Mannino has expressed it:

> What's beyond my comprehension is why turning rocks into happiness elsewhere should matter at all. That strikes me as okay, but still utterly useless and therefore immoral if it comes at the opportunity cost of not preventing suffering. The non-creation of happiness is not problematic, for it never results in a problem for anyone (i.e. any consciousness-moment), and so there's never a problem you can point to in the world; the non-prevention of suffering, on the other hand, results in a problem.[41]

Philosopher David Benatar further argues that we not only should avoid bringing (overtly) miserable lives into existence, but that we ideally should avoid bringing *any* lives into existence, since coming into existence is *always* a harm on his account. Benatar's main argument rests on the premise that the absence of suffering is good, while the absence of happiness is not bad, and hence the state of non-existence is good ("good" + "not bad" = "good"). In contrast, the presence of suffering and happiness is bad and good respectively, and hence not a pure good, which renders it worse than the state of non-existence, according to Benatar.[42]

Beyond this value asymmetry, Benatar further argues that there is an asymmetry in how much suffering and happiness our lives contain — e.g. the worst forms of suffering are far worse than the best pleasures are good; we almost always experience some subtle unpleasantness, dissatisfaction, and preference frustration; and there are such negative things as chronic pain, impairment, and trauma, yet no such thing as chronic pleasure. And the reason that we fail to acknowledge this, Benatar argues, is that we have various psychological biases that cause us to evaluate our lives in overly optimistic terms.[43]

It seems worth expanding a bit on this more quantitative asymmetry between the respective badness and goodness of suffering and happiness. For even if one rejects the notion that there is an asymmetry between the moral status of creating happiness and preventing suffering for future beings, it seems difficult to deny Benatar's claim that the worst forms of suffering are far worse

---

40 Singer, 1980. However, Singer goes on to say about this view of coming into existence that it "perhaps, is a reason to combine [preference and hedonistic utilitarianism]". Furthermore, Singer has since moved closer toward, and now defends, hedonistic utilitarianism, whereas he was arguably primarily a preference utilitarian when he made the quoted statement. (Peter Singer has confirmed this, personal communication.)
41 Quoted from a Facebook conversation.
42 Benatar, 2006, ch. 2.
43 Benatar, 2006, ch. 3; see also Vinding, 2020a, ch. 7.

than the best of pleasures are good. Imagine, for example, that we were offered ten years of the greatest happiness possible on the condition that we must undergo some amount of hellish torture in order to get it. How much torture would we be willing to endure in order to get this prize? Many of us would reject the offer completely and prefer an entirely unproblematic state over any mixture of hellish torture and heavenly happiness.[44]

Others, however, will be willing to accept the offer and make a sacrifice. Yet how big a sacrifice could one reasonably be willing to make to gain ten years of maximal bliss? Seconds of hellish torture? A full hour? Perhaps even an entire day? Some might go as far as saying an entire day, yet it seems that very few would push the scale to anywhere near 50/50.[45]

Those who would be willing to endure a full day of torture in order to enjoy ten years of paradise are, I think, among those who are willing to push it the furthest in order to attain such happiness.[46] Yet notice how far they are from 50/50. We are not talking 80/20, 90/10, or even 99/1 here. No, one day of hell for 3650 days of paradise roughly corresponds to a "days of happiness to days of suffering" ratio of 99.97 to 0.03. And that is for those who are willing to push it.

So not only is there no symmetry here, but the moral weight of the worst of suffering appears to be orders of magnitude greater than that of the greatest happiness, which suggests that the prevention of suffering should be our main priority on any plausible moral calculus. Even on a view according to which we are willing to really push it and endure an unusually large amount of suffering, the vast majority of moral weight is *still* found in preventing suffering, at least when speaking in terms of durations of the best and worst potential states. And one can reasonably argue that this is also true of the actual state of the world, as Arthur Schopenhauer did when comparing "the feelings of an animal engaged in eating another with those of the animal being eaten."[47]

A more general and qualitative asymmetry between the moral status of happiness and suffering has been defended by philosopher Karl Popper:

> I believe that there is, from the ethical point of view, no symmetry between suffering and happiness, or between pain and pleasure. … In my opinion human suffering makes a

---

44 In one survey (n = 99), almost half of the respondents said that they would not endure one minute of intense suffering for any amount of happiness for themselves, Tomasik, 2015c.
45 Cf. Tomasik, 2015c.
46 Cf. Tomasik, 2015c.
47 Schopenhauer, 1851/1970, p. 42; see also Vinding, 2020a, sec. 1.2.

direct moral appeal, namely, the appeal for help, while there is no similar call to increase the happiness of a man who is doing well anyway. A further criticism of the Utilitarian formula "Maximize pleasure" is that it assumes a continuous pleasure-pain scale which allows us to treat degrees of pain as negative degrees of pleasure. But, from the moral point of view, pain cannot be outweighed by pleasure, and especially not one man's pain by another man's pleasure. Instead of the greatest happiness for the greatest number, one should demand, more modestly, the least amount of avoidable suffering for all; …[48]

In other words, suffering carries an inherent urgency, whereas the creation of happiness does not. We would rightly rush to send an ambulance to help someone who is enduring extreme suffering, yet not to boost the happiness of someone who is already happy, no matter how much we may be able to boost such a well-off person's feelings of joy. Similarly, if we had pills that could raise the happiness of those who are already doing well to the greatest heights possible, there would be no urgency in distributing these pills (to those already doing well), whereas if a single person fell to the ground in unbearable agony right before us, there would indeed be an urgency to help. Increasing the happiness of the already happy is, unlike the alleviation of extreme suffering, not an emergency.

A similar consideration about the Abolitionist Project championed by David Pearce — i.e. the abolition of suffering throughout the living world via biotechnology — lends further credence to this asymmetrical view of the moral status of creating happiness versus preventing suffering.[49] Imagine that we had completed the Abolitionist Project and made suffering non-existent for good in all lifeforms, such that all sentient beings only experience gradients of bliss. Would our moral obligations be exactly the same after this completion as before? Would we have an equally strong duty to move sentience to new heights of pleasure after we had abolished suffering? Or would we instead have discharged our prime moral obligation, and thus have reason to lower our shoulders and breathe a deep and justified sigh of moral relief? I think the latter.

Another reason in favor of an asymmetrical view is that, echoing Benatar, it seems that the absence of extreme happiness is not bad in remotely the same way as the presence of extreme suffering is bad. For example, if a person is in a state of dreamless sleep rather than having the experience of a lifetime, this cannot reasonably be characterized as a disaster or a catastrophe; the difference between these two states does not seem to carry great moral weight, if any. Yet when it comes to the difference between sleeping and being tortured, we are indeed talking about

---

48 Popper, 1945/2011, note 2, ch. 9.
49 Pearce, 2007.

a difference that does carry immense moral weight, and the realization of the worse rather than the better outcome would clearly amount to a catastrophe.

The final asymmetry I shall review in this section is one that is found more on a meta-level, namely in the distribution of views concerning the moral value of the creation of happiness versus the prevention of suffering. For in our broader human conversation about what has value, very few seem to have seriously disputed the disvalue of suffering and the importance of preventing it. Indeed, to the extent that we can find a value that almost everyone agrees on, it is this: suffering matters. In contrast, there are many who have disputed the value and importance of creating more happiness. These include many thinkers in Eastern philosophy for whom moksha — liberation from suffering — is the highest good, as well as many thinkers in Western philosophy, including Epicurus and Arthur Schopenhauer.

This asymmetry in consensus about the value and moral status of creating happiness versus preventing suffering also counts as a weak reason for giving greater priority to the latter.[50]

6.2 Tranquilism: Happiness as the Absence of Suffering

Author Lukas Gloor defends a view he calls tranquilism, which holds that the value of happiness lies in its absence of suffering.[51] Thus, according to tranquilism, states of euphoric bliss are not of greater value than, say, states of peaceful contentment free of any negative components.

Tranquilism is well in line with the asymmetry in moral status between happiness and suffering. And one may even argue that it explains this asymmetry: if the value of happiness lies chiefly in its absence of suffering, then it follows that creating happiness (for the untroubled) cannot take precedence over reducing suffering. Moving someone from one unproblematic state to another can never constitute a greater move on the value scale than moving someone from a problematic state to a less problematic one.

To many of us, this is a highly counterintuitive view, at least at first sight. After all, do we not seek pleasure almost all the time, often at the seemingly justified cost of suffering? Yet one can frame this seeking in another way that is consistent with tranquilism, by viewing our search for pleasure as really being an attempt to escape suffering and dissatisfaction. On this framing, what appears to be a transition from neutral to positive is really a transition from a state of negativity, however subtle, to a state that is relieved, at least to some extent, from this negativity.

---

50 For more on this asymmetry in consensus, see Vinding, 2020a, sec. 1.5.
51 Gloor, 2017.

So, on this view, when we visit a friend we have desired to see for some time, we do not go from a neutral to a positive state, but instead just remove our craving for their company and the dissatisfaction caused by their absence. The same goes for the pleasure of physical exercise: it is liberating in that it gives us temporary freedom from the bad feelings and moods that follow from not exercising. Or even the pleasure of falling in love, which provides refreshing relief from the boredom and desire we are otherwise plagued by.

Psychologist William James seemed to agree with this view of happiness:

> Happiness, I have lately discovered, is no positive feeling, but a negative condition of freedom from a number of restrictive sensations of which our organism usually seems the seat. When they are wiped out, the clearness and cleanness of the contrast is happiness. This is why anaesthetics make us so happy.[52]

Arthur Schopenhauer also endorsed this view.[53] Here is how Lukas Gloor explains it:

> In the context of everyday life, there are almost always things that ever so slightly bother us. Uncomfortable pressure in the shoes, thirst, hunger, headaches, boredom, itches, non-effortless work, worries, longing for better times. When our brain is flooded with pleasure, we temporarily become unaware of all the negative ingredients of our stream of consciousness, and they thus cease to exist. Pleasure is the typical way in which our minds experience temporary freedom from suffering, which may contribute to the view that happiness is the symmetrical counterpart to suffering, and that pleasure, at the expense of all other possible states, is intrinsically important and worth bringing about.[54]

One may object that the implication that mere contentment has the same value as the greatest euphoric bliss seems implausible, and thus counts against tranquilism. Yet whether this is implausible depends on the eyes that look. After all, if someone is experiencing "mere contentment" without any negative cravings whatsoever, and if this person does not find their experience insufficient in any way, who are we to say that they are wrong about their state, and that they actually should want something better? Tranquilism denies that such a "merely content"

---

52 James, 1901.
53 Schopenhauer, 1851/1970, p. 41.
54 Gloor, 2017.

person is wrong to claim that their state is perfect. Indeed, tranquilism is in complete agreement with this person's own assessment, and hence this implication of tranquilism is at least not implausible from that person's perspective, which one may argue is the most relevant perspective to consider in this context. Moreover, a proponent of tranquilism may argue that the implication only appears implausible from the perspective of someone who is *not* in a state of perfect contentment — one who desires to see more euphoric bliss in the world, and who in some sense feels lacking, i.e. a painful craving, about its absence.

The implications of tranquilism are clear: creating more happiness (for the currently non-existent or otherwise untroubled) has neutral value, while there *is* value in the alleviation and prevention of suffering.

6.3 Creating Happiness at the Price of Suffering Is Wrong

In this section I shall not argue for a novel, separate point, but instead invoke some concrete examples that help make the case for a claim that follows directly from some of the views that we have seen above, the claim being that it is wrong to create happiness at the price of suffering.

One obvious example of such gratuitous suffering would be that of torturing a single person for the enjoyment of a large crowd.[55] If we think happiness can always outweigh suffering, we seem forced to say that, yes, provided that the resulting enjoyment of the crowd is great enough — and if other things are equal — then such happiness can indeed outweigh and justify torturing a single person. Yet that seems misguided. A similar example to consider is that of a gang rape: if we think happiness can always outweigh suffering, then such a rape can in principle be justified, provided that the pleasure of the rapists is sufficiently great. Yet most people would find this proposition utterly wrong.[56]

One may object that these thought experiments bring other issues into play than merely that of happiness versus suffering, which is a fair point. Yet we can in a sense control for such extraneous factors by reversing the purpose of these acts so that they are about reducing suffering rather than increasing happiness for a given group of individuals. So rather than the torture of a single person being done for the enjoyment of a crowd, it is now done in order to prevent a crowd from being tortured in even worse ways; rather than the rape being done for the pleasure of, say, ten people, it is done to prevent ten people from being raped in even more agonizing

---

55 There are various versions of this example. A common one is whether it can be right to make gladiators fight for the enjoyment of a full colosseum, which is often raised as a problematic question for (certain versions of) utilitarianism.
56 This example is due to Richard Ryder, Ryder, 2011, ch. 3.

ways. While we may still find it most unpalatable to allow such preventive actions, it nonetheless seems plausible that torturing a single person in order to prevent many people from experiencing more severe torture would be the right thing to do, and that opting for less rape to occur is better than opting for more instances of even more agonizing rape to occur.

A similar example is the situation described in Ursula K. Le Guin's short story *The Ones Who Walk Away from Omelas*. The story is about an almost utopian city, Omelas, in which everyone lives an extraordinarily happy and meaningful life, except for a single child who is locked in a basement room, fated to live a life of squalor:

> The child used to scream for help at night, and cry a good deal, but now it only makes a kind of whining, "eh-haa, eh-haa," and it speaks less and less often. It is so thin there are no calves to its legs; its belly protrudes; it lives on a half-bowl of corn meal and grease a day. It is naked. Its buttocks and thighs are a mass of festered sores, as it sits in its own excrement continually.[57]

The story ends by describing some people in the city who appear to find the situation unacceptable and who choose not to take part in it any more — the ones who walk away from Omelas.

The relevant question for us to consider is whether we would walk away from Omelas, or rather whether we would choose to bring a condition like Omelas into existence in the first place. Can the happy and meaningful lives of the other inhabitants of Omelas justify the existence of this single, miserable child? Different people have different intuitions about it. Some will say that it depends on how many people live in Omelas. Yet for many of us, the answer is "no" — the creation of happiness is comparatively frivolous and unnecessary, and it cannot justify the creation of such a victim and such misery and suffering.[58]

A "no" to the creation of Omelas would also be supported by the Asymmetry in population ethics, according to which it has neutral value to add a happy life to Omelas, while adding this one miserable child has negative value, and hence the net value of the creation of Omelas is negative.

---

57 Le Guin, 1973/1992.

58 And even though many people may insist that the child's suffering is a worthy sacrifice, the fact that it only takes a single life of misery to bring the value of a whole paradisiacal city into serious question is yet another strong hint that there is an asymmetry between the (dis)value of happiness and suffering.

6.4 Brief Reply to a Common Objection

A common objection against suffering-focused views is that they imply that it would be right to kill everyone (or "destroy the world"). One reply to this claim is that many suffering-focused views do not carry this implication. For example, in his book *The Battle for Compassion: Ethics in an Apathetic Universe*, Jonathan Leighton argues for a pragmatic position he calls "negative utilitarianism plus", according to which we should aim to do our best to reduce preventable suffering, yet where we can still "categorically refuse to intentionally destroy the planet and eliminate ourselves and everything we care about in the process".[59]

Another reply is that alternative ethical views that have a strong consequentialist component seem about as vulnerable to similar objections.[60] For instance, if maximizing the sum of happiness minus suffering were our core objective, one could think that we ought to kill people in order to replace them with happier beings. It may then be objected, quite reasonably, that this is unlikely to be optimal in practice. Yet one can argue — just as reasonably, I believe — that the same holds true of trying to destroy the world in order to reduce suffering: it is not the best we can do in practice.[61]

Having visited this general case for suffering-focused ethics, we shall now turn to what I consider the strongest case for such a view.[62]

---

59 Leighton, 2011, p. 96; see also Mayerfeld, 1999, p. 159.
60 Knutsson, 2021.
61 For a more elaborate discussion, see Vinding, 2020a, sec. 8.1; Knutsson, 2021; Ajantaival, 2022.
62 Additional arguments for suffering-focused ethics are found in Vinding, 2020a, Part I.

7. The Principle of Sympathy for Intense Suffering

No positive good can outweigh the very worst forms of suffering. This common view lies at the heart of the ethical principle I shall argue for in this chapter. What this principle says, roughly speaking, is that we should prioritize the interests of those who experience the most extreme forms of suffering. In particular, we should prioritize their interest in not experiencing such suffering higher than we should prioritize anything else. I call this the "principle of sympathy for intense suffering".[63]

7.1 Motivation and Outline

Something that may help motivate us to take this principle seriously is to ask, quite simply, why we would suppose it to be otherwise. Why should we believe that the most extreme forms of suffering can somehow be outweighed or counterbalanced by something else? After all, there seems no reason, *a priori*, to suppose this to be the case. That is, before we look at the particulars, if we consider only what we know about the world at large, we have little reason to assume that the most extreme suffering should be "outweighable". By all appearances, the world was not designed to make the well-being of sentient beings "net positive". Nor does anything we know about the universe at large guarantee that such "net positive" well-being is even a possibility in principle. In short, we have no *a priori* reason to assume a positive story in this regard.

It may be objected that we have no reason to assume a negative story either. And this is true. The point here is just that we should make no strong *a priori* assumptions in either direction. What I shall argue, however, is that when we proceed to look at the relevant particulars — especially the felt experience of extreme suffering — we do indeed, as sympathetic beings, find overwhelming reason to draw a negative conclusion.

A first glance at the particulars may appear to suggest the opposite conclusion. After all, common sense would seem to say that if a conscious subject considers some state of suffering worth experiencing in order to attain some given pleasure, then this pleasure is indeed worth the

---

63 I do not claim that this principle is original to me. For example, it bears strong similarity to what Brian Tomasik calls "consent-based negative utilitarianism", Tomasik, 2015a, as well as to the ethical view defended by Jonathan Leighton (Leighton, 2011). An essentially identical view is also explored in Gloor, 2016, sec. II. Jamie Mayerfeld, too, has expressed a view similar to the one I present in this chapter, Mayerfeld, 1999, p. 148, p. 178.

suffering. Yet this view runs into serious problems in cases where subjects consider their suffering unoutweighable by any amount of pleasure.

For what would the common-sense view say in such a situation? That the suffering indeed cannot be outweighed by any pleasure? That would seem an intuitive suggestion. Yet the problem is that we can also imagine the case of an experience of some pleasure that the subject, in that experience-moment, considers so great that it can outweigh even the worst forms of suffering, which leaves us with mutually incompatible value claims (although one can reasonably doubt the existence of such positive states, whereas, as we shall see below, the existence of correspondingly negative experiences is a certainty). How are we to evaluate these claims?

The common-sense method of evaluation invoked above has clearly broken down at this point, and is entirely silent on the matter. We are forced to appeal to another principle of evaluation. And the principle we should employ is, I maintain, to sympathize with those who are the worst off. Hence the principle of sympathy for intense suffering: we should sympathize with the evaluations of those subjects who experience suffering so intense that they 1) consider it unbearable — i.e. they cannot consent to it even if they try their hardest — and 2) consider it unoutweighable by any positive good, even if only for a brief experience-moment. More precisely, we should minimize the amount of such experience-moments of extreme suffering.[64]

7.2 Common-Sense Support

This principle has a lot of support from common sense. For example, imagine that two children are offered to ride a roller coaster — one child would find the ride very pleasant, while the other would find it very *un*pleasant. And imagine, furthermore, that the only two options available are that they either both ride or neither of them ride (and if neither of them ride, they are both perfectly fine). Which of these options should we ideally realize?

Common sense would seem to say that we should sympathize with and prioritize the child who would find the ride very unpleasant, and hence choose the outcome in which there is no harm and no victim. The mere pleasure of the "ride-positive" child does not justify a violation of the interest of the other child not to suffer a very unpleasant experience. The interest in not enduring such suffering is far more fundamental, and hence has ethical primacy. Indeed, one can argue that there is no genuine interest to have a very pleasant experience if the absence of such an

---

64 Note that, as I defend this principle, it applies both intrapersonally and interpersonally, yet one may also accept it only in the interpersonal case (which is the most morally relevant case at any rate).

experience is also considered perfectly okay and unproblematic (of course, children in the real world who want roller coaster rides rarely consider it wholly unproblematic if they do not ride).

Arguably, common sense even suggests the same in the case where there are many more children who would find the ride very pleasant, while still only one child who would find it very unpleasant (provided, again, that the children will all be perfectly fine if they do not ride). And yet the suffering in this example, a very unpleasant experience on a roller coaster, can hardly be said to count as extreme, much less an instance of the worst forms of suffering — the forms of suffering that constitute the strongest support for the principle of sympathy for intense suffering. Such intense suffering, even if balanced against the most intense forms of pleasure, only demands even stronger relative priority. However bad we may consider the imposition of a very unpleasant experience for the sake of a very pleasant one, the imposition of extreme suffering for the sake of extreme pleasure must be deemed far worse.

7.3 The Horrendous Support

The worst forms of suffering are so terrible that merely thinking about them for a brief moment can leave us in a state of horror for a good while. Thus, we naturally prefer not to contemplate these things. Yet if we are to make sure that we have our priorities right, and that our views about what matters most in this world are as well-considered as possible, then we cannot shy away from the task of contemplating and trying to appreciate the disvalue of these worst of horrors. This is no easy task, and not just because we are reluctant to think about the issue in the first place, but also because it is difficult to gain anything close to a true appreciation of the reality in question. As David Pearce writes:

> It's easy to convince oneself that things can't *really* be that bad, that the horror invoked is being overblown, that what is going on elsewhere in space-time is somehow less real than *this* here-and-now, or that the good in the world somehow offsets the bad. Yet however vividly one thinks one can imagine what agony, torture or suicidal despair must be like, the reality is inconceivably worse. Hazy images of Orwell's 'Room 101' barely hint at what I'm talking about. The force of 'inconceivably' is itself largely inconceivable here.[65]

---

65 Pearce, 1995, 2.7. A personal anecdote of mine in support of Pearce's quote is that, although I write a lot about reducing suffering, I am *always* unpleasantly surprised by how bad it is when I experience even just borderline intense suffering. I then always get the sense that I have no idea what I am talking about when I discuss suffering in my usual happy state, although the words I use in that state are quite accurate: that it is *really* bad. It truly is inconceivable, as we simply cannot simulate that badness in a remotely faithful way when we are feeling good, analogous to the phenomenon of binocular rivalry, where we can only perceive one of two visual images at a time.

Many torture victims have echoed this sentiment of inconceivability and inexpressibility. For example, Belgian resistance worker Jean Améry, who was caught and tortured by the Nazis, reported that nothing could have prepared him for the first blow inflicted on him by the Gestapo, as if it revealed a whole new, previously inconceivable dimension of existence:

> One may have known about torture and death in the cell [before being tortured], without such knowledge having possessed the hue of life … But with the first blow … a part of our life ends and it can never again be revived.[66]

A similar report was given by journalist Jacobo Timerman, who was captured and tortured by the Argentine military regime of the late 1970s:

> In the long months of confinement, I often thought of how to transmit the pain that a tortured person undergoes. And always I concluded that it was impossible. It is a pain without points of reference, revelatory symbols, or clues to serve as indicators.[67]

These sobering remarks notwithstanding, we can still gain at least some (admittedly very limited) appreciation of the horror of extreme suffering by considering some real-world examples of it. (What follows are examples of an extremely unpleasant character that may be triggering and traumatizing.)

We can begin to approach such appreciation by considering what it is like to undergo "death by sawing", an execution method used on humans in many parts of the world in the past, and which is still used on non-human animals in some parts of the world today.[68] The victim is tied up and then sawn or chopped in half with a saw or a machete, either transversely or midsagittally — in the latter case, it has both been done from the groin up as well as from the skull down.[69]

---

66 The quote is from Améry's *At the Mind's Limits: Contemplations by a Survivor on Auschwitz*, as quoted in Mayerfeld, 1999, p. 38.
67 The quote is from Timerman's *Prisoner Without a Name, Cell Without a Number*, as quoted in Mayerfeld, 1999, p. 42.
68 See e.g. Baker, 2015.
69 See e.g. Busnot, 1714, ch. 8. Beyond such intentionally caused deaths, similar deaths in which the victim is cut or torn apart also happen without human intention, such as due to traffic and industrial accidents, as well as predation.

A similarly excruciating way to die is to be tortured to death inside a brazen bull: a hollow metal bull designed to contain a person inside while a fire is burning underneath, which slowly roasts the person inside to death. Many people have been reported to die in this way in ancient times.[70] And many more beings have died, and continue to die, in similar ways, such as by being burned or boiled alive, which happens countless times every day. Indeed, the latter fate — being boiled to death — happens to roughly a million non-human beings each year in the slaughterhouses of the United States alone.[71]

Another horrifying example of torturous suffering is what happened to Kuwaiti Ahmad Qabazard, who, while only a teenager, was captured by Iraqi soldiers and tortured brutally for weeks because of his involvement with the Kuwaiti resistance movement. The soldiers sought to make Ahmad reveal the names of fellow resistance fighters, yet he refused to do so. Finally, Ahmad's parents were told that he would soon be released. They heard a car approaching and went to the door:

> When Ahmad was taken out of the car they saw that his ears, his nose and his genitalia had been cut off. He was coming out of the car with his eyes in his hands. Then the Iraqis shot him, once in the stomach and once in the head, and told his mother to be sure not to move the body for three days.[72]

No less horrible is the tragic fate of the Japanese girl Junko Furuta, who was kidnapped by four teenage boys in 1988, at the age of 16. According to their own trial statements, the boys raped her hundreds of times; inserted scissors and skewers into her vagina and anus; beat her with golf clubs, bamboo sticks, and iron rods; tied her hands to the ceiling and used her as a punching bag; set fireworks into her anus, vagina, and ear; burned her vagina with cigarettes; tore off her left nipple with pliers; and more. Eventually, she was no longer able to move from the ground, and she repeatedly begged the boys to kill her, which they eventually did, after 44 days, by pouring lighter fluid on her face and body, and setting her on fire.[73]

---

70 For instance, the Romans are reported to have killed Christians with the brazen bull, and the bull is also reported to have been used on overthrown tyrants, such as Burdunellus, who was reportedly killed in this way in 497, Collins, 2004, p. 35.
71 Dockterman, 2013.
72 Report by Julie Flint, *Observer*, March 3, 1991, as quoted in Glover, 2000, p. 32.
73 The official trial documents for the case (in Japanese, yet copyable for machine translation) can be found at: courts.go.jp/app/files/hanrei_jp/261/020261_hanrei.pdf

Another example of extreme suffering is what happened to airplane pilot Dax Cowart. In 1973, at the age of 25, Dax went on a trip with his father to visit land that he considered buying. Unfortunately, due to a pipeline leak, the air over the land was filled with propane gas, which is highly flammable when combined with oxygen. As they tried to start their car, the propane ignited, and the two men found themselves in a burning inferno. Dax's father died, and Dax himself had much of his hands, eyes, and ears burned away; two thirds of his skin was severely burned.

The case of Dax has since become quite famous, not only because of the extreme horror he experienced during and after the explosion, but also because of the ethical issues raised by his treatment, which turned out to be about as torturous as the explosion itself. For Dax himself repeatedly said, immediately after the explosion as well as for months later, that he wanted to die more than anything else, and that he did not want to be subjected to any treatment that would keep him alive. Nonetheless, he was forcibly treated for a period of 14 months, during which he tried to end his life several times.

After his treatment, Dax managed to recover and live what he considered a happy life — he successfully sued the oil company responsible for the pipeline leak, which left him financially secure; he earned a law degree; got married; and reported to have had "some very, very good experiences" after the accident.[74] Yet even from this position of an accomplished and self-reportedly happy life, he still wished that he had been euthanized rather than treated. In Dax's own view, no happiness could ever compensate for what he went through.[75]

This kind of evaluation is exactly what the ethical principle advocated here centers on, and what the principle amounts to is essentially a refusal to claim that Dax's evaluation, or any other like it, is wrong. The principle of sympathy for intense suffering maintains that we should not allow the occurrence of such unbearable suffering that the subject finds unoutweighable by any positive good.

And it is indeed worth reflecting on what a rejection of this view would entail. For if one holds that extreme suffering *can* be outweighed by some positive good, one ought to clarify how. Specifically, according to whom, and measured by what criteria, is such suffering outweighable by positive goods? The most promising option that lies open in this regard, it seems, is to prioritize the assessments of beings who say that their happiness, or other good things about their

74 Engel, 1983.
75 Engel, 1983; Benatar, 2006, p. 63.

lives, do outweigh and justify such extreme suffering endured by others.[76] Yet that, I maintain, would be a profoundly unsympathetic choice.

I shall spare the reader from further examples of extreme suffering here in the text, and instead refer to sources, found in the following footnote, that contain additional cases that are worth considering in order to gain an appreciation of the moral significance of extreme suffering.[77] And the crucial question we must ask ourselves in relation to these examples is whether the creation of happiness or any other purported good could ever justify the creation of, or the failure to prevent, suffering this bad and worse. If not, this implies that our priority should not be to create happiness or other alleged goods, but instead to prevent extreme suffering, regardless of where in time and space it may risk emerging.

7.4 Objection: What About That Which Is Most Precious?

Among the objections against this view I can think of, the strongest, at least at first sight, is the sentiment: but what about that which is most precious in your life? What about the person who is most dear to you? If anything stands a chance of outweighing the disvalue of extreme suffering, surely this is it. In more specific terms: does it not seem plausible to claim that saving the most precious person in one's life could be worth an instance of the very worst suffering?

Yet one has to be careful about how this question is construed. If what we mean by "saving" is that we save them from extreme suffering, then we are measuring extreme suffering against extreme suffering, and hence we have not pointed to a distinct value entity that might outweigh the disvalue of extreme suffering. Therefore, if we are to point to such a distinct entity, "saving" must here mean something that does not itself involve extreme suffering. And if we wish to claim that there is something wholly different from the reduction of suffering that can be put on the scale, the fate we are saving this person from (at the supposedly worthwhile price of extreme suffering) should preferably involve no suffering at all. So what we should consider is a choice between 1) the mixed bargain of an instance of the very worst of suffering, i.e. unendurable suffering, and the continued existence of the most precious person one knows, or 2) the painless discontinuation of the existence of this person, yet without any ensuing suffering for others or oneself.

---

76 After all, if such happy beings themselves deny that the good things in their lives can outweigh extreme suffering, it hardly seems plausible for us to say that they do.
77 reducing-suffering.org/the-horror-of-suffering
reducing-suffering.org/on-the-seriousness-of-suffering
simonknutsson.com/the-seriousness-of-suffering-supplement
youtube.com/watch?v=RyA_eF7W02s ["Preventing Extreme Suffering Has Moral Priority"]

Now, when phrased in this way, choosing 1) may not sound all that bad to us, especially if we do not know the one who will suffer. Yet this would be cheating — nothing but an appeal to our all too partial moral intuitions. It clearly betrays the principle of impartiality, according to which it should not matter whom the suffering in question is imposed upon; it should be considered equally disvaluable regardless. Thus, if we abide by the principle of impartiality, we may equivalently phrase the choice above as being between 1) the continued existence of the most precious person one knows, yet at the price that this person has to experience a state of unendurable suffering — a state they themself would consider so bad that, according to them, it could never be outweighed by any intrinsic good — or 2) the discontinuation of the existence of this being without any ensuing suffering.

When phrased in this way, it actually seems clearer to me than ever that 2) is the superior choice, and that we should adopt the principle of sympathy for intense suffering. For how could one possibly justify imposing such extreme and, according to the subject, unoutweighable suffering upon the most precious person one knows, suffering that this person would, in that moment, rather die than continue to experience? In this way, for me at least, it is no overstatement to say that this objection against the principle of sympathy for intense suffering, when considered more carefully, actually ends up being one of the strongest arguments in its favor.[78]

The principle of sympathy for intense suffering defended here stems neither from depression nor resentment. Rather, as the name implies, it simply stems from a deep sympathy for intense suffering. It stems from a firm choice to side with the evaluations of those who are superlatively worst off. And while it is true that this principle has the implication that it would have been better if the world had never existed, I think the fault here is to be found in the world, not the principle.

Some pockets of the universe are in a state of insufferable darkness. Such suffering is like a black hole that sucks all light out of the world. Or rather, if all the light of the world has any intrinsic value, it pales in comparison to the disvalue of this darkness. Yet, by extension, this also implies that there *is* a form of light whose value *does* compare to this darkness, and that is the kind of light we should aspire to become: the light that brightens and prevents the unendurable darkness of the world.[79]

---

78 Replies to additional objections against the principle of sympathy for intense suffering can be found in Vinding, 2020a, sec. 4.5 and sec. 8.10-8.12.

79 I suspect that both the content and phrasing of the last couple of sentences are inspired by the following quote I saw written on Facebook by Robert Daoust: "What is at the center of the universe of ethics, I suggest, is not the sun of the good and its play of bad shadows, but the black hole of suffering."

8. Anti-Speciesism

— Peter Singer, *The Effective Altruism Handbook*

Speciesism, as noted earlier, is discrimination against beings based on their species membership.[81] If we are to stay true to the core virtue of impartiality, then such discrimination must be rejected.

8.1 The Indefensibility of Our Prevailing View

The prevailing view of the respective moral status of human and non-human individuals is that the former matter far more than the latter, if the latter matter at all. The question, then, is what reasons can be given for this difference in our view. A common response is that the reason for our discrimination is that humans are more intelligent than other animals. Yet the problem with this claim is that we do not assign differential moral value to human individuals on the basis of their "intelligence", no matter how broadly we construe this term. Humans who do not possess certain cognitive abilities, such as the ability to speak or plan for the future, are not considered less morally valuable than those who do possess these abilities, which forces us to reconsider this purported reason for our discrimination. As Peter Singer notes: "If possessing a higher degree of intelligence does not entitle one human to use another for his or her own ends, how can it entitle humans to exploit nonhumans for the same purpose?"[82]

So if not intelligence, what might then be the reason for our discrimination? A more sophisticated argument might be that the crucial difference is that humans are more sentient than other animals. Because we have a larger brain, we feel pain and suffering much more intensely than do members of other species, and therefore our discrimination is justified.

Yet the claim that humans are more sentient than other animals is quite dubious. For on what grounds can we assert that, say, a mouse or a fish feels pain any less intensely than does a

---

80 Carey, 2015, p. 98.
81 Horta, 2010.
82 Carey, 2015, p. 98.

human? The mere fact that they have smaller brains is not by itself a compelling reason. After all, children have smaller brains than adults, yet should children feel pain any less intensely than adults? It appears that they do not, and the opposite could well be the case. The truth is that we have little reason to believe that a larger brain, be it in terms of its number of neurons or in terms of its mass, necessarily gives rise to more intense experiences.

Sure, it seems reasonable to expect that there is *some* relationship between brain complexity and the quality and intensity of the experiences a brain can give rise to.[83] Yet this does not imply that bigger brains necessarily give rise to more intense experiences than do smaller brains, just as we cannot say that beings with bigger thighs necessarily run faster than beings with smaller thighs, even though there surely is *some* relationship between thigh size and running ability. Humans, for instance, have larger thighs than do domestic cats, but that clearly does not imply that we can outrun them. Similarly, it might well be the case that our larger brains do not mediate more intense experiences than do the brains of, say, domestic cats, or any other animal with the relevant brain structures. Indeed, as Steven Pinker notes, "We have every reason to believe that cats suffer as much as humans do."[84]

Our direct investigations into the physical signatures of conscious experience likewise give us reason to doubt that the number of neurons found in the brain is of the greatest relevance when it comes to the intensity of its experiences. For instance, we know that more than half of the neurons in the human brain are found in the cerebellum, a brain structure that does not seem to play an important role for our conscious experience, if any, while a minority of neurons are found in the cerebrum and the limbic system, which appear to mediate most if not all aspects of our conscious experience. Hence, what seems most relevant to the capacity to experience conscious states such as pain and suffering is not the number of neurons found in the brain, but rather certain brain structures, and these structures are found across the vertebrate line, with analogous structures found in many invertebrates. They are anything but uniquely human.[85]

And not only may other beings experience suffering as intensely as humans do, but they may even experience it more intensely. As David Pearce notes:

---

83 Oscar Horta comments (edited slightly): "There is certainly a relationship between brain complexity and the complexity of experiences, but regarding the intensity... It's a common claim, yet it is just a supposition. I haven't seen any evidence for it. The intensity of a bad experience is different from how nuanced that experience is."
84 Quoted from Pinker, 2012.
85 See e.g. the Cambridge Declaration on Consciousness, which notes that sentience might well have arisen long before vertebrate species arose, implying that sentience may be shared not only by all vertebrates, but also by many invertebrates too, Low et al., 2012.

We often find it convenient to act as though the capacity to suffer were somehow inseparably bound up with linguistic ability or ratiocinative prowess. Yet there is absolutely no evidence that this is the case, and a great deal that it isn't. The functional regions of the brain which subserve physical agony, the "pain centres", and the mainly limbic substrates of emotion, appear in phylogenetic terms to be remarkably constant in the vertebrate line. … Not merely is the biochemistry of suffering disturbingly similar where not effectively type-identical across a wide spectrum of vertebrate (and even some invertebrate) species. It is at least possible that members of any species whose members have more pain cells exhibiting greater synaptic density than humans sometimes suffer more atrociously than we do, whatever their notional "intelligence".[86]

The possibility that beings of other species may experience suffering more intensely than humans has also been raised by zoologist James Serpell, who notes that, unlike other animals, humans can rationalize their pain, which may help reduce it.[87] Richard Dawkins has made an evolutionary argument for the same conclusion:

I can see a Darwinian reason why there might even be a negative correlation between intellect and susceptibility to pain. …

Isn't it plausible that a clever species such as our own might need less pain, precisely because we are capable of intelligently working out what is good for us, and what damaging events we should avoid? Isn't it plausible that an unintelligent species might need a massive wallop of pain, to drive home a lesson that we can learn with less powerful inducement?

At very least, I conclude that we have no general reason to think that non-human animals feel pain less acutely than we do, and we should in any case give them the benefit of the doubt.[88]

But even if we were to grant that the smaller brains of most non-human animals mediate less intense experiences than do the bigger brains of humans, this would still not justify disregarding them in moral terms. After all, there are also humans who have much smaller brains than most

---

86 Pearce, 1995, sec. 1.9.
87 See about ten minutes into the movie *Speciesism: The Movie*.
88 Dawkins, 2011. See also the YouTube video "Richard Dawkins: No Civilized Person Accepts Slavery, So Why Do We Accept Animal Cruelty?"

other humans, such as those who are born without cerebral hemispheres.[89] Of course, when it comes to such humans, it is clear that the assumption that smaller-brained beings suffer less than bigger-brained ones is highly questionable, and the opposite cannot readily be ruled out — i.e. they may well experience many sensations more intensely than do humans with bigger brains. Yet even if we were to grant that humans with substantially smaller brains suffer less, what would the implications be? Not that significant, it seems, as it is clear that all sentient humans matter morally regardless of the size of their brains. It is clear that they are beings whose suffering we have good reason to alleviate, and whose fundamental interests should not be trumped by the trivial interests of others. And, needless to say, the same applies to all sentient beings whose brains are similarly different from the average human brain. To believe otherwise would be to engage in discrimination based on species membership.

More generally, for any cognitive ability we may point to as the crucial ability that non-human animals lack (and which supposedly justifies discrimination against them), we can likewise point to human individuals who do not possess this ability. Yet when it comes to such human individuals, very few seem tempted to argue that we are justified in discriminating against them based on such differences. Indeed, we tend to actively deny that this is the case, which suggests that these differences are not the real reason we discriminate against non-human individuals, and that we instead merely invoke these differences as a rationalization.

As the above indicates, the root source of our speciesism is not a careful examination of the neurophysiology of non-human animals, but instead, rather unsurprisingly, something much more crude. In particular, our moral intuitions appear to be animated by a "bodyism" of sorts. After all, we would never find it justifiable to morally disregard and kill a sentient being with a human body just because this being has the mind of a cow or a chicken, or any other kind of mind. So why, then, do we find it acceptable to morally disregard and kill a cow or a chicken with reference to the fact that they have the mind of a cow or a chicken? Clearly, the mind is not the determining factor for us. If only we transplanted the mind-brain of any sentient non-human animal into a human body, this would apparently be enough to protect this being from our speciesism. Yet this perfectly exposes our ethical failure on this point, as it obviously betrays one of the most fundamental ethical values we all hold, namely that we should not value a being differently based on the external appearances of that being.

---

89 For instance, some children with hydranencephaly have no cerebral cortex at all, Beshkar, 2008.

Quite simply, our speciesist attitudes and practices cannot be justified.[90] Discrimination based on species membership is no more justifiable than discrimination against humans based on their race, sex, or cognitive abilities, and it should be condemned just as strongly. Add to this the fact that non-human beings constitute the vast majority of sentient beings on the planet — more than 99.99 percent of sentient vertebrates are non-human[91] — and it becomes apparent just how skewed and indefensible our prevailing moral and altruistic priorities are.

One might object that our focus is not skewed if we count the total number of brain cells found in different classes of animals. Yet this is not true, since there are more than a hundred times as many neurons in the total mass of insect brains alone than there are in all human brains, and there likewise seem to be around 10 to 100 times more fish neurons than there are human neurons.[92] What if we also factor in brain-to-body mass ratio? This only favors insects more, as they generally have large such ratios, much larger than those of humans. Small ants, for example, have a brain-to-body mass ratio of 1:7, while it is around 1:40 in humans. Indeed, several small mammals and birds have a higher brain-to-body mass ratio than humans, as does at least one species of fish.[93]

However much we may try, there is just no way to avoid the conclusion that our prevailing altruistic and moral focus is indefensibly narrow. And as long as that remains true — as long as virtually all of our altruistic attention is devoted to much less than 0.01 percent of the sentient beings on the planet — our attempts at being *effectively* altruistic are bound to fail.

8.2 The Implications of Rejecting Speciesism

The main implications of rejecting speciesism are quite straightforward, if difficult to come to terms with for minds as anthropocentrically schooled as ours. One implication is that we should stop imposing gratuitous suffering on non-human beings. In particular, we should abolish animal agriculture and end our practice of consuming and otherwise frivolously exploiting them. For just as a rejection of racism is not compatible with keeping other humans as property, much less eating and otherwise exploiting their body parts, a rejection of speciesism is not compatible with commodifying, eating, and frivolously exploiting individuals belonging to other species.

Sadly, we are far detached from our own core principles on this inconvenient issue. Not only are we discriminating against beings based on their external appearances, but we also fail to stay true

---

90 For a more elaborate case for this conclusion, see Vinding, 2015b; Horta, 2022.
91 Tomasik, 2009c.
92 Ray, 2017,
93 See e.g. Nilsson, 1996, pp. 603-604. As for whether insects feel pain, see Tomasik, 2009a; Singer, 2016.

to the concern for suffering that most of us claim to have. After all, many people say that they will not accept the imposition of torturous suffering on any individual for any reason, least of all for the sake of pleasure. Yet the consumption choices of most people fly directly in the face of this claim.

For example, immense amounts of intense suffering are caused by the fishing industry. Fish are the vertebrates that humanity exploits and kills in the greatest numbers by far, usually without even the pretense of doing it "humanely" — they are torn out of the ocean, suffocated, and decapitated without regard for the extreme pain and suffering caused in the process.[94]

Our actions are thus starkly inconsistent with our proclaimed opposition to torture and torturous suffering. Contra high-minded ideals of walking away from Omelas, we find ourselves actively walking right into the opposite scenario, a world in which countless beings live like the miserable child in Omelas — compare that child's life with the average life of a pig on a factory farm — and many of these beings are routinely subjected to horrors that are far worse still, all for the sake of the pleasure and convenience of a small minority of beings on the planet. Yet as George Bernard Shaw observed, "custom will reconcile people to any atrocity".[95]

8.3 Wild Animal Suffering

Ceasing to actively harm non-human animals for morally frivolous reasons is far from the only practical implication that follows from a rejection of speciesism. We also have strong reasons to start helping them. After all, the vast majority of sentient beings on the planet — and hence the vast majority of suffering by virtually any measure — is found not in farms, but in the wild. In terms of presently existing beings, this is where the greatest scope for altruism is found.[96]

This may be surprising to many people. After all, aren't non-human animals in nature mostly well off? And why should we help them rather than just let them be? The former sentiment reflects a common idyllic view of nature, a view that is sadly wrong.[97] As Peter Singer notes: "Evolution is an impersonal natural process that has no regard for the wellbeing of the individual creatures it has produced."[98] And the reality of life in the wild reflects this fact all too well: it is a condition of abject poverty.

---

94 On the sentience of fish, see Balcombe, 2016.
95 Salt, 1915, preface.
96 Tomasik, 2009c.
97 See e.g. Faria & Paez, 2015.
98 Singer, 2016.

Just like humans who live in extreme poverty, non-human beings in nature suffer from readily preventable diseases, starvation, parasitism, lack of basic safety, etc.[99] And just like humans, non-human individuals have an interest in avoiding such misery, which leads us to the second question: why should we care and do something about it? In short, because not considering the interests of these beings would be an indefensible, speciesist position. When it comes to humans, we realize that we have strong reasons to do something to alleviate conditions of abject poverty and extreme misery, even if we have played no part in creating these ills. And to think that this changes when the afflicted beings are non-human individuals simply amounts to unjustified discrimination. It is to fail to give proper consideration to the interests of beings merely because they belong to another species.[100]

But what can we do about such suffering? This is no doubt a difficult question to answer in much detail at this early stage. Some answers have been proposed, however, one of them being the practice of "compassionate biology" defended by David Pearce, which entails cross-species fertility-regulation and the use of modern technology to create a "pan-species welfare state in tomorrow's Nature reserves: in short, 'high-tech Jainism'."[101]

Yet even if we had no proposed answers of this sort, we would still not have reason to dismiss the issue altogether. Again, consider the case of human poverty: the mere fact that this is a difficult problem to solve does not mean that we should give up and do nothing. On the contrary, it implies that we should do more research and think harder about how we might be able to improve the situation. So too with the problem of suffering in the wild: rather than ignore it or give up, we have good reasons to do more research. Supporting such research efforts is one thing we can do at this point. Another is to promote the idea that we should reduce suffering in nature, and to promote concern for non-human individuals in general, both of which are the focus of the organization Animal Ethics. Thus, the idea that there is nothing we can do about this problem at this point is really quite false.[102]

8.4 We Should Expect to Be Extremely Biased About Speciesism

Our speciesist bias has already been alluded to in an earlier chapter, yet it is still worth delving deeper into it here. For if ever there were a bias that we evolved not to transcend, it is surely this. After all, we evolved in a context in which our survival depended on our killing and eating non-

99 Tomasik, 2009b; Faria & Paez, 2015.
100 On why most people do not seem to care about wild animal suffering, see Davidow, 2013; Vinding, 2020b.
101 Pearce, 2016.
102 Animal Ethics, 2013.

human beings. For most of our evolutionary history, the questioning of such a practice, and the belief that non-human beings should be taken seriously in moral terms, meant a radically decreased probability of survival and reproduction. And this would likely also apply to one's entire tribe if one were to start spreading such a sentiment, which might help explain the visceral threat that many people seem to feel upon engaging with supporters of this sentiment today. In other words, having significant moral concern for non-human beings was not a recipe for survival in our evolutionary past. It was more like a death sentence. For this reason alone, we should expect to be extremely biased on this matter.

And yet this evolutionary tale is far from the full story, as there is also a cultural story to be told, which provides even more reasons to expect our outlook to be intensely biased. For on top of our (likely) speciesist biological hardware, we also have the software of cultural programming running, and it runs the ultimate propaganda campaign against concern for non-human beings. Indeed, if we ran just a remotely similar campaign against humans, we would consider the term "propaganda" a gross understatement. Their body parts are for sale at every supermarket and on the menu for virtually every meal; their skin is ripped off their bodies and used for handbags, shoes, and sports equipment; their names are used pejoratively in every other sentence. Why, indeed, would anyone expect this to leave our moral cognition with respect to these beings biased in the least? Or rather, why should we expect to stand any chance whatsoever of having just a single rational thought about the moral status of these beings? Well, we arguably shouldn't — not without immense amounts of effort spent rebelling against our crude nature and the atrocious culture that it has spawned.

Another bias that is relevant to consider, on top of the preceding considerations, is that human altruism tends to be motivated by a hidden drive to show others how cool and likable we are, and to increase our own social status.[103] To think that we transcend this motive merely by calling ourselves "effective altruists" would be naive. The problem, then, is that rejecting speciesism and taking the implications of such a rejection seriously is, sadly, seen as quite uncool at this point. If one were to do so, one would become more than a little obnoxious and unlikeable in the eyes of most people, and be more like a favorite object of ridicule than of admiration, none of which is enticing for social creatures like us. So even if reason unanimously says that we should reject speciesism, we have a thousand and one social reasons that say just the opposite.

As mentioned earlier, there are also psychological studies that demonstrate the existence of strong biases in our views of non-human individuals, such as that we "value individuals of certain species less than others even when beliefs about intelligence and sentience are accounted

103 Cf. Simler & Hanson, 2018.

for".[104] More than that, we diminish and even deny the mental capacities of the kinds of beings whom we consider food — a denial that is increased by "expectations regarding the immediate consumption" of such beings.[105]

These forms of bias should give us pause and should encourage serious reflection. The way forward, it seems, is to admit that we are extremely biased and to commit to doing better.

---

104 Caviola et al., 2019, abstract.
105 Bastian et al., 2012, abstract.

9. Future Directions

"We have enormous opportunity to reduce suffering on behalf of sentient creatures."

— Brian Tomasik[106]

So far we have seen a general introduction to some of the main ideas associated with effective altruism, followed by a brief case for suffering-focused ethics and a critique of our anthropocentric betrayal of the principle of impartiality.

Given all of this, what are the implications? How do we best help others? Or to phrase this question in more specific terms in line with the values I have argued for: how can we best reduce extreme suffering for all sentient beings?

The first thing we should remind ourselves upon considering this question is, once again, that it is an open question — a significant amount of uncertainty regarding how we can best act in accordance with our values is unavoidable. With this in mind, let us see what answers we might reasonably propose.[107]

9.1 Promoting Deeper Concern for Suffering

One endeavor that seems particularly promising, and which I have attempted to pursue in this book, is that of promoting a deeper conversation about values, especially with an emphasis on the moral significance of extreme suffering. It is likely that the change in values toward greater concern for extreme suffering that could be furthered via such a conversation is among the most important changes we can make. The main reason for this, in short, is that our values are a key driver of our actions.[108]

People who believe this to be a promising strategy include Brian Tomasik, who has presented a case for it in his essay "Reasons to Promote Suffering-Focused Ethics", and Simon Knutsson, who has made questions concerning fundamental values, and suffering-focused ethics in

---

106 Tomasik, 2013a.
107 A deeper exploration of how we can best reduce suffering is found in Vinding, 2020a, Part II.
108 I have made a more elaborate case in favor of this strategy in Vinding, 2020a, ch. 12. For a case for focusing on the reduction of suffering in politics, along with an examination of what politics aimed at the reduction of suffering might look like, see Vinding, 2022.

particular, the main focus of his work.[109] As for how such deeper reflection and concern can best be fostered, it seems promising to formulate and spread clear presentations of arguments in favor of suffering-focused ethics, as well as to familiarize people with some of the unimaginably horrible real-life cases of extreme suffering.[110]

To get a clearer sense of why a focus on values is important, and how divergent the implications of different values can be, consider the practical implications of the following two moral principles: 1) we will not allow the creation of a single instance of the worst forms of suffering for any amount of happiness, and 2) we will allow one day of the worst suffering for ten years of the most sublime happiness. What kind of future would we accept with these respective principles?

Imagine a future in which we colonize space and maximize the number of sentient beings that the accessible universe can sustain over the entire course of the future. This number is probably more than 10^{30}.[111] Given this number of beings, and assuming that these beings each live a hundred years, principle 2) above would permit a colonization of space that creates more than 10^{28} years of the most intense suffering, provided that the other states of experience are sublimely happy. This is how extreme the practical difference can be between whether we consider suffering irredeemable or not. And notice that even if we altered the exchange rate by orders of magnitude — say, by requiring 10^{15} times more sublime happiness per unit of extreme suffering than we did in principle 2) above — we would still allow an enormous amount of extreme suffering to be created; in the concrete case of requiring 10^{15} times more happiness, we would allow more than 10,000 billion years of the worst forms of suffering.

9.2 Reducing S-Risks

The preceding discussion brings us to another strategy, namely to reduce risks of astronomical future suffering, also referred to as "s-risks".[112] For as much as we may hope that suffering will never occur at an astronomical scale, there is still a risk that it might.

It can be difficult to see the relevance of this. After all, how could things possibly turn out so badly? Yet one can say a couple of things in response to such a sentiment. First, the probability

---

109 Tomasik, 2015b; Knutsson, 2016.
110 Additional proposals are found in Vinding, 2020a, sec. 12.5.
111 Indeed, Nick Bostrom puts the "lower bound of the number of biological human life-years in the future accessible universe" at 10^{34}, cf. Bostrom, 2013, p. 18, and in his book *Superintelligence*, he estimates that, given that human consciousness can be emulated in different substrates, the accessible universe could sustain 10^{60} such life-years, cf. Bostrom 2014, p. 103.
112 Althaus & Gloor, 2016.

of a very bad outcome need not be large for the expected value of trying to prevent it to be high. Second, the probability of an outcome that involves suffering on an astronomical scale is arguably not that small when we consider the many different ways such an outcome could come about.[113] Third, we know that we have a tendency to engage in wishful thinking, which should make us skeptical of our inclination to dismiss these risks of bad outcomes.[114]

So how may we best prevent s-risks? Some believe the optimal way is to influence the development of emerging technologies, especially artificial intelligence. Other promising strategies include the one outlined above, promoting deeper concern for suffering, as well as the one described in the following section: expanding our moral circle.[115]

9.3 Expanding Humanity's Moral Circle

An endeavor that seems promising for reducing future suffering is to expand humanity's circle of moral consideration. Indeed, it seems that increased concern for suffering and moral circle expansion are both necessary yet insufficient for effectively reducing future suffering. For if we only expand our moral circle, yet fail to deepen our concern for extreme suffering, we will not reduce the risk of scenarios of the kind outlined above, with astronomical amounts of extreme suffering allowed for the sake of creating large amounts of happiness. At the same time, if we only deepen our concern for suffering without also expanding our moral circle, such deepened concern may end up not encompassing all beings who can suffer.

One way to see the importance of expanding the moral circle is to consider how humanity currently devotes almost all its altruistic resources to a tiny fraction of sentient beings on the planet, while neglecting, and in many cases actively causing, the suffering of the rest. Additionally, there is a risk that other kinds of vulnerable non-human beings will emerge in the future, such as humanly created (i.e. "artificial") non-human beings of some kind. Indeed, not only might such non-human beings eventually outnumber humans as much as non-human animals do today, but they may even come to outnumber humans (or other agents in power) by orders of magnitude more.

How the moral circle of the future can best be expanded stands as an open question. Yet it seems that the promotion of concern for the morally excluded non-human beings who currently exist is a reasonable place to start, whether done by exposing people to the extensive cruelty that is being

---

113 Tomasik, 2011; Althaus & Gloor, 2016; Baumann, 2017.
114 See also Tomasik, 2013a.
115 See also Vinding, 2020a, ch. 14; Baumann, 2021.

inflicted upon non-human animals, or by presenting arguments against speciesism.[116] Such increased concern would not only reduce a lot of suffering in the near term, but would also help us think more clearly about non-human beings in general, which may put us on a better trajectory for the future.

9.4 Researching and Reflecting on the Question

The question of how we can best reduce extreme suffering is not merely open, but also highly neglected, which makes further research on it seem of great value. And there are indeed many diverse questions that are relevant to explore in this context, from how a moral priority to extreme suffering can best be promoted, to what institutions are most conducive to the prevention of extreme suffering. These questions matter deeply for our endeavor to reduce suffering, as the sophistication of our answers to them largely determines how effectively we can pursue this endeavor.[117]

Research on these questions is pursued at various organizations associated with effective altruism, including the Center on Long-Term Risk and the Center for Reducing Suffering (again, I should note that I am a co-founder of CRS). Both of these organizations have an elaborate research agenda that can be found on their respective websites.[118]

9.5 Far From Omelas

I should like to conclude this chapter by re-emphasizing a tragic fact that is all too easily forgotten by our wishful and optimistic minds, that fact being that the world we inhabit is hopelessly far from Omelas. For our world is unfortunately nothing like a near-paradisiacal city predicated on the misery of a single child. Rather, in our world, there are millions of starving children, and millions of children who die from such starvation or otherwise readily preventable causes, every single year. And none of this misery serves to support a paradise or anything close to it.

We do not live in a world where a starving child confined to a basement is anywhere near the worst forms of suffering that exist. Sadly, our world contains an incomprehensibly larger number of horrors of incomprehensibly greater severity, forms of suffering that make the sufferer wish dearly for a fate as "lucky" as that of the unfortunate child in Omelas. This is, of course, true

---

116 Cf. Vinding, 2016b.
117 For more on the research project necessary for reducing suffering, see Vinding, 2020a, ch. 16.
118 See longtermrisk.org/research-agenda; centerforreducingsuffering.org/open-research-questions

even if we only consider the human realm, yet it is even more true if we also, as we must, consider the realm of non-human individuals.

Humanity subjects billions of land-living beings to conditions similar to those of the child in Omelas, and we inflict extreme suffering upon a significant fraction of them, by castrating them without anesthetics, boiling them alive, suffocating them, grinding them alive, etc. And our sins toward aquatic animals are greater still, as we kill them in *far* greater numbers, trillions on some estimates; and most tragically, these deaths probably involve extreme suffering more often than not, as we slowly drag these beings out of the deep, suffocate them, and cut off their heads without stunning or mercy. And yet even this horror story of unfathomable scale still falls hopelessly short of capturing the true extent of suffering in the world, as the suffering created by humanity only represents a fraction of the totality of suffering on the planet. The vast majority of this suffering is found in the wild, where non-human animals suffer and die from starvation, parasitism, and disease, not to mention being eaten alive, which is a source of extreme suffering for countless beings on the planet every single second.

Sadly, our condition is very far from Omelas, implying that if one would choose to walk away from Omelas, it seems impossible to defend supporting the spread of our condition, or anything remotely like it, beyond Earth. The extent of suffering in the world is immense and overshadowing, and our future priorities should reflect this reality.

10. A Healthy Life: The Precondition for Effective Altruism

In this final chapter, I will briefly address a tempting pitfall related to effective altruism, namely that of failing to invest sufficiently in oneself. For as noted in the introduction, effective altruism is not altruism in the irresponsible sense of sacrificing for others without any regard for oneself. Not only would that amount to an indefensible neglect of an individual of great inherent value, i.e. the altruist, but it is also just an ineffective and unsustainable strategy for helping others. In other words, effective altruism necessitates self-care of the highest quality.

If one believes that one is able to make a significant difference in the world, then one must also believe that one is worth the substantial level of self-investment that is required to realize one's altruistic potential. Indeed, self-investments might well be among the best investments one can make as an altruist. After all, the impact of most altruistic investments tends to be highly uncertain, whereas many of us can be quite sure that our future selves will be dedicated to the mission of helping others (given our track record, pledges, peers, etc.).[119]

With extreme suffering crying out for alleviation every single second, it may seem trivial and out of place to suggest that the very best thing one can do is to brush one's teeth carefully and to change one's sheets. Yet, as counter-intuitive as such a suggestion is, it is quite likely true, at least occasionally. Real-life effective altruism is not akin to heroically jumping into a pond to save a child once in a lifetime (although it may include that), but rather a matter of living a healthy and productive life guided by reflection and evidence, over the long haul.

Consider it this way: of all the altruistic projects you can undertake in your lifetime, how likely is it that the one you are currently working on is the single most important one? Or more to the point, how likely is it that it is more important than all the other projects you can do for the rest of your life combined? Unlikely, it seems safe to say. The importance of the project that one happens to focus on at any given moment, especially at an early stage in one's career, is all but surely dwarfed by the importance of all future projects combined. This suggests that quality self-care is generally at least as important as the particular altruistic project that one is currently working on. In other words, minimizing "personal existential risks" and taking good care of ourselves in a long-term oriented fashion should be a prime priority.

---

119 I am grateful to Rubí Amoros for helping me see this point.

10.1 How Can We Best Help Ourselves?

So how is this best done? We probably all know the most basic answers, which are those boring things that wise people have been saying since the dawn of time: eat well, sleep well, exercise well, and limit the time you spend on social media. These basic things are all crucial, and they are probably among the first things to focus on if we are to have a healthy lifestyle.

Yet important as those basics are, they are clearly far from being sufficient for living a healthy and impactful life of compassionate purpose.[120] If we are to advance beyond those core basics, while still keeping things simple, it seems to me that the Hindu concept of Purusartha provides a helpful framework.[121] Purusartha means "object of human pursuit", and refers to what is considered the four proper goals of human life in the Hindu tradition, the necessary and sufficient goals for a good and fulfilling life.

The first of these goals is Dharma, which is basically about being a good person, but in all senses of the word. It is not enough to merely donate money to effective charities or to do altruistically important work. One must also be a kind and respectful person, to others and to oneself — a person who seeks to live virtuously and in accordance with the local duties and laws necessary for social life to function. It is about living up to one's ethical responsibility in the broadest sense. And this includes having the decency and compassion to not succumb to "effective altruist snobbery" in the form of looking down upon or mocking supposedly ineffective ways of helping, or deriding problems or misery because they may not seem to be of the most urgent kind. No matter how much extreme suffering there is in the world, it does not change the intrinsic badness of the distress of, say, a single homeless person, nor the fact that their distress is tragic and lamentable. To scorn at suffering and squalor, or at genuine attempts to help alleviate such misfortune, is not effective altruism, but merely a display of a lack of respect and empathetic understanding.

Simply put, Dharma is the embodiment of the decency, responsibility, and kindness that not only makes life better for others around us, but which also tends to be deeply satisfying to ourselves. For example, giving to and doing things for others has been found to promote the wellbeing of the benefactor too.[122] But it is also satisfying in that striving to live a responsible and ethical life can give us the feeling that we are living in a socially sustainable way that is in alignment with our own deepest values and highest purpose.

---

120 Cf. Vinding, forthcoming.
121 I am not claiming that this framework is all that unique; one can readily find similar ideas in other traditions, as well as in recent books on positive psychology.
122 Dunn et al., 2008; Dunn et al., 2010; Mogensen, 2011.

Second, there is Artha, which is about having the basic means of life in order. It is about having a satisfying career and financial security, which enables one to live with a sense of safety, direction, and opportunities. It is also about having one's home, belongings, and daily rhythm and routine in order, keeping things clean and functional.

Yet being an ethical person with a good career and an orderly home is not enough, of course. We humans are social animals who have deep needs beyond merely being paragons of virtue and career success. We also have needs for social and sensual connection, and this is the dimension of life covered by the concept of Kama (as in "Kama Sutra"). Like Dharma, Kama is also a very broad notion, one that stands for pleasure, but of many sorts. It is about desire, lust, and pleasures of the senses, but it is also about the deep loving connections we have with those closest to us. It is about having kindred spirits with whom we can share things openly and honestly, and who can give us comfort and support.

Lastly, there is Moksha, which means liberation from suffering. In some schools of Hinduism, this concept has strong superstitious connotations, but in others, as well as in the secular interpretation proposed here, it simply refers to the self-knowledge and freedom from suffering that one can attain via spiritual practice, where "spiritual" is just a synonym for "meditative" or "contemplative". Such freedom may be brought about via yoga, meditation, or a quiet walk.

In other words, Moksha, as interpreted here, basically refers to a tranquilist cleaning practice of detaching from mental noise. For many of us, such a practice is crucial for living with less stress and anxiety, as well as for living with more resilience and peace. Indeed, as a large meta-analysis of studies on mindfulness meditation found: "[Mindfulness-based therapy] is an effective treatment for a variety of psychological problems, and is especially effective for reducing anxiety, depression, and stress."[123]

Each of the four elements listed above seem, if not necessary, then at least helpful for most of us in order to live a healthy life in the sense most worth striving for — a life in which we work for the benefit of all sentient beings, and in which we are empowered by whatever connections and restorative practices that we need in order to be fulfilled and effective in the world. In short, a life in which we strive to be the light that brightens the darkness of the world.

---

123 Khoury et al., 2013.

Acknowledgments

Bibliography

Ajantaival, T. (2022). Peacefulness, nonviolence, and experientialist minimalism. https://centerforreducingsuffering.org/research/peacefulness-nonviolence-and-experientialist-minimalism/

Althaus, D. & Gloor, L. (2016). Reducing Risks of Astronomical Suffering: A Neglected Priority. https://longtermrisk.org/reducing-risks-of-astronomical-suffering-a-neglected-priority/

Animal Charity Evaluators. (2018). Donation Impact. https://web.archive.org/web/20181116210226/https://animalcharityevaluators.org/donation-advice/donation-impact/

Animal Ethics. (2012/2016). Speciesism. http://www.animal-ethics.org/speciesism/

Animal Ethics. (2013/2016). Working for a future with fewer harms to wild animals. http://www.animal-ethics.org/wild-animal-suffering-section/helping-animals-in-the-wild/working-for-a-future-with-fewer-harms-to-wild-animals/

Animal Ethics. (2017). Scope insensitivity: Failing to appreciate the numbers of those who need our help. https://www.animal-ethics.org/scope-insensitivity-failing-to-appreciate-the-numbers-of-those-who-need-our-help/

Baker, K. (2015). Pigs chopped in half with ceremonial machetes while they're still alive… so villagers can dip money in their blood for good luck: Vietnamese ceremony condemned by animal groups. https://www.dailymail.co.uk/news/article-2966668/Pigs-chopped-half-ceremonial-machetes-alive.html

Balcombe, J. (2016). *What a Fish Knows: The Inner Lives of Our Underwater Cousins*. Farrar, Straus and Giroux.

Bastardi, A. et al. (2011). Wishful Thinking: Belief, Desire, and the Motivated Evaluation of Scientific Evidence. *Psychological Science*, 22(6), pp. 731-732.

Bastian, B. et al. (2012). Don't Mind Meat? The Denial of Mind to Animals Used for Human Consumption. *Personality and Social Psychology Bulletin*, 38(2), pp. 247-256.

Baumann, T. (2017). S-risks: An introduction. http://s-risks.org/intro/

Baumann, T. (2021). How can we reduce s-risks?
https://centerforreducingsuffering.org/research/how-can-we-reduce-s-risks/

Baumann, T. (2022). Career advice for reducing suffering.
https://centerforreducingsuffering.org/research/career-advice-for-reducing-suffering/

Benatar, D. (2006). *Better Never to Have Been: The Harm of Coming into Existence*. Oxford University Press.

Bertrand, O. (2016). Reducing suffering or adding years of life? Effective altruism and divergences in value judgments and worldviews. Presented at the 14th Conference of the International Society for Utilitarian Studies.
https://www.academia.edu/26705132/Reducing_suffering_or_adding_years_of_life_Effective_altruism_and_divergences_in_value_judgments_and_worldviews

Beshkar, M. (2008). The presence of consciousness in the absence of the cerebral cortex. *Synapse*, 62(7), pp. 553-556.

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

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.

Busnot, D. (1714/2015). *History of the Reign of Muley Ismael, the Present King of Morocco*. Ishi Press International.

Carey, R. (2015). *The Effective Altruism Handbook*. Centre For Effective Altruism.

Caviola, L. et al. (2019). The Moral Standing of Animals: Towards a Psychology of Speciesism. *Journal of Personality and Social Psychology*, 116(6), pp. 1011-1029.

Collins, R. (2004). *Visigothic Spain, 409-711*. Blackwell Pub.

Davidow, B. (2013). Why Most People Don't Care About Wild-Animal Suffering.
http://reducing-suffering.org/why-most-people-dont-care-about-wild-animal-suffering/

Dawkins, R. (2011). Richard Dawkins on vivisection: "But can they suffer?".
http://boingboing.net/2011/06/30/richard-dawkins-on-v.html

Dockterman, E. (2013). Nearly One Million Chickens and Turkeys Unintentionally Boiled Alive Each Year in U.S. Time Magazine.
http://nation.time.com/2013/10/29/nearly-one-million-chickens-and-turkeys-unintentionally-boiled-alive-each-year-in-u-s/

Dunn, E. et al. (2008). Spending Money on Others Promotes Happiness. *Science*, 319(5870), pp. 1687-1688.

Dunn, E. et al. (2011). If money doesn't make you happy, then you probably aren't spending it right. *Journal of Consumer Psychology*, 21(2), pp. 115-125.

Engel, M. (1983). A Happy Life Afterward Doesn't Make Up for Torture. https://www.washingtonpost.com/archive/opinions/1983/06/26/a-happy-life-afterward-doesnt-make-up-for-torture/ab680b30-237b-4b7a-b6da-1f7ab3da9208/

Faria, C. & Paez, E. (2015). Animals in need: The problem of wild animal suffering and intervention in nature. *Relations: Beyond Anthropocentrism*, 3, pp. 7-13.

Fehige, C. (1998). A pareto principle for possible people. In Fehige, C. & Wessels U. (eds.), *Preferences*. Walter de Gruyter.

GiveWell. (2012/2022). GiveWell's Cost-Effectiveness Analyses. https://www.givewell.org/how-we-work/our-criteria/cost-effectiveness/cost-effectiveness-models

GiveWell. (2016/2021). Process for Identifying Top Charities. https://www.givewell.org/how-we-work/process

GiveWell. (2021/2022). How We Produce Impact Estimates. https://www.givewell.org/impact-estimates

Gloor, L. (2016/2019). The Case for Suffering-Focused Ethics. https://longtermrisk.org/the-case-for-suffering-focused-ethics/

Gloor, L. (2017). Tranquilism. https://longtermrisk.org/tranquilism/

Glover, J. (2000). *Humanity: A Moral History of the Twentieth Century*. Yale University Press.

Hanson, R. (2014). I Still Don't Get FOOM. https://www.overcomingbias.com/2014/07/30855.html

Horta, O. (2010). What Is Speciesism? *Journal of Agricultural and Environmental Ethics*, 23, pp. 243-266.

Horta, O. (2022). *Making a Stand for Animals*. Routledge.

James W. (1901). Letter on happiness to Miss Frances R. Morse. In James H. (ed.), *Letters of William James. Vol. 2*. Atlantic Monthly Press, 1920.

Karnofsky, H. (2011). Why we can't take expected value estimates literally (even when they're unbiased). https://blog.givewell.org/2011/08/18/why-we-cant-take-expected-value-estimates-literally-even-when-theyre-unbiased/

Karnofsky, H. (2014). Sequence thinking vs. cluster thinking. https://blog.givewell.org/2014/06/10/sequence-thinking-vs-cluster-thinking/

Khoury, B. et al. (2013). Mindfulness-based therapy: A comprehensive meta-analysis. *Clinical Psychology Review,* 33(6) pp. 763-771.

Knutsson, S. (2015/2016). The Seriousness of Suffering: Supplement. http://www.simonknutsson.com/the-seriousness-of-suffering-supplement

Knutsson, S. (2016). Effective Altruism and Ethics.
https://www.simonknutsson.com/effective-altruism-and-ethics

Knutsson, S. (2017). Simon Knutsson – Suffering-Focused Ethics and Effective Altruism.
https://www.youtube.com/watch?v=2CfBcHii06w

Knutsson, S. (2021). The World Destruction Argument. *Inquiry*, 61(10), pp. 1004-1023.

Le Guin, U. (1973/1993). *The Ones Who Walk Away from Omelas*. Creative Education.

Leighton, J. (2011). *The Battle for Compassion: Ethics in an Apathetic Universe*. Algora Pub.

Low, P. et al. (2012). The Cambridge Declaration on Consciousness.
http://fcmconference.org/img/CambridgeDeclarationOnConsciousness.pdf

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

MacAskill, W. (2017). Effective Altruism: Introduction. *Essays in Philosophy*, 18(1), pp. 1-5.

Mayerfeld, J. (1999). *Suffering and Moral Responsibility*. Oxford University Press.

Mogensen, A. (2011). Giving Without Sacrifice? The relationship between income, happiness, and giving.
https://assets.ctfassets.net/dhpcfh1bs3p6/3jqEdXNDmg8SGAeOeC8KYo/63f9caa4de752b15c84455dc9025d537/giving-without-sacrifice.pdf

Moor, M. (2013). Impact through rationality.
https://www.youtube.com/watch?v=PcWus1943K0

Narveson, J. (1973). Moral problems of population. *The Monist*, 57(1), pp. 62-86.

Nilsson, G. (1996). Brain and body oxygen requirements of *Gnathonemus petersii*, a fish with an exceptionally large brain. *J Exp Biol*, 199(3), pp. 603-607.

Pearce, D. (1995/2007). *The Hedonistic Imperative*. Published online at:
http://www.hedweb.com/hedab.htm

Pearce, D. (2007). The Abolitionist Project.
https://www.abolitionist.com/

Pearce, D. (2016). Compassionate Biology: How CRISPR-based "gene drives" could cheaply, rapidly and sustainably reduce suffering throughout the living world.
https://gene-drives.com/

Pearce, D. (2017). *Can Biotechnology Abolish Suffering?* The Neuroethics Foundation.

Pinker, S. (2002). *The Blank Slate: The Modern Denial of Human Nature*. Penguin.

Pinker, S. (2012). The Great Debate.
https://www.youtube.com/watch?v=ooTcyioNIZ4&ab_channel=PGPlaysVideoGames

Plous, S. (1993). *The Psychology of Judgment and Decision Making*. McGraw-Hill.

Popper, K. (1945/2011). *The Open Society and Its Enemies*. Routledge.

Ray, G. (2017/2018). How many neurons are there?
https://eukaryotewritesblog.com/how-many-neurons-are-there/

Ryder, R. (2011/2017). *Speciesism, Painism and Happiness: A Morality for the 21st Century*. Andrews UK Ltd.

Salt, H. (ed.) (1915). *Killing for Sport*. G. Bell and Sons.

Schopenhauer, A. (1851/1970). *Essays and Aphorisms*. Penguin Books.

Sidgwick, H. (1874/1981). *The Methods of Ethics*. Hackett Pub. Co.

Simler, K. & Hanson, R. (2018). *The Elephant in the Brain: Hidden Motives in Everyday Life*. Oxford University Press.

Singer, P. (1972). Famine, Affluence, and Morality. *Philosophy and Public Affairs*, 1(1), pp. 229-243.

Singer, P. (1980). Right to Life?
http://www.nybooks.com/articles/1980/08/14/right-to-life/

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

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

Singer, P. (2016). Are Insects Conscious?
https://www.project-syndicate.org/commentary/are-insects-conscious-by-peter-singer-2016-05?barrier=accessreg

Sinick, J. (2013). Some reservations about Singer's child-in-the-pond argument.
http://lesswrong.com/lw/hr5/some_reservations_about_singers_childinthepond/

Slovic, P. (2007). "If I look at the mass I will never act": Psychic numbing and genocide. *Judgment and Decision Making*, 2(2), pp. 79-95.

Speciesism: The Movie. (2012). Film. Directed by Mark Devries. Mark Devries Productions.

Tenney, E. et al. (2016). Does positivity enhance work performance? Why, when, and what we don't know. *Research in Organizational Behavior*, 36, pp. 27-46.

Todd, B. (2015). Why you should focus more on talent gaps, not funding gaps.
https://80000hours.org/2015/11/why-you-should-focus-more-on-talent-gaps-not-funding-gaps/

Tomasik, B. (2006/2016). On the Seriousness of Suffering.
http://reducing-suffering.org/on-the-seriousness-of-suffering/

Tomasik, B. (2007/2016). Why Maximize Expected Value?
http://reducing-suffering.org/why-maximize-expected-value/

Tomasik, B. (2009a/2014). Do Bugs Feel Pain?
http://reducing-suffering.org/do-bugs-feel-pain/

Tomasik, B. (2009b/2020). The Importance of Wild-Animal Suffering.
https://longtermrisk.org/the-importance-of-wild-animal-suffering/

Tomasik, B. (2009c/2014). How Many Wild Animals Are There?
http://reducing-suffering.org/how-many-wild-animals-are-there/

Tomasik, B. (2011/2016). Risks of Astronomical Future Suffering.
https://longtermrisk.org/risks-of-astronomical-future-suffering/

Tomasik, B. (2013a/2015). Against Wishful Thinking.
https://longtermrisk.org/against-wishful-thinking/

Tomasik, B. (2013b/2016). The Horror of Suffering.
http://reducing-suffering.org/the-horror-of-suffering/

Tomasik, B. (2015a). Are Happiness and Suffering Symmetric?
https://reducing-suffering.org/happiness-suffering-symmetric/

Tomasik, B. (2015b). Reasons to Promote Suffering-Focused Ethics.
http://reducing-suffering.org/the-case-for-promoting-suffering-focused-ethics/

Tomasik, B. (2015c). A Small Mechanical Turk Survey on Ethics and Animal Welfare.
https://reducing-suffering.org/a-small-mechanical-turk-survey-on-ethics-and-animal-welfare/

Tomasik, B. (2016). Preventing Extreme Suffering Has Moral Priority [graphic content in middle of video].
https://www.youtube.com/watch?v=RyA_eF7W02s

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

Vinding, M. (2015). *Speciesism: Why It Is Wrong and the Implications of Rejecting It.*

Vinding, M. (2016a). *Reflections on Intelligence*.

Vinding, M. (2016b/2020). Animal advocates should focus on antispeciesism, not veganism.
https://magnusvinding.com/2020/05/04/animal-advocates-should-focus-on-antispeciesism-not-veganism/

Vinding, M. (2017). A Contra AI FOOM Reading List.
https://magnusvinding.com/2017/12/16/a-contra-ai-foom-reading-list/

Vinding, M. (2020a). *Suffering-Focused Ethics: Defense and Implications*. Ratio Ethica.

Vinding, M. (2020b). Ten Biases Against Prioritizing Wild-Animal Suffering.
https://magnusvinding.com/2020/07/02/ten-biases-against-prioritizing-wild-animal-suffering/

Vinding, M. (2022). *Reasoned Politics*. Ratio Ethica.

Vinding, M. (forthcoming). *Compassionate Purpose: Personal Inspiration for a Better World.* Ratio Ethica.

WHO. (2020). Children: Improving survival and well-being. https://www.who.int/en/news-room/fact-sheets/detail/children-reducing-mortality

Wiblin, R. (2017). What are the most important talent gaps in the effective altruism community? https://80000hours.org/2017/11/talent-gaps-survey-2017/