# The Psychology of Effective Altruism

**Course Syllabus**

## Tuesdays 4:00 to 6:30 pm

##### Dane Smith Hall 226

**Spring Term 2020**

**Psych 450 section 5, CRN 35810**

## Instructor:

## Geoffrey Miller, Associate Professor of Psychology

Office hours: Tuesdays 1:30 – 2:30 pm, Logan Hall 160

gfmiller@unm.edu, 277-1967 (office)

**Brief course description**

This seminar will cover the psychological challenges of how we can help others more effectively, given the limited resources and trade-offs of a complicated world. It is centered around a new moral movement called ‘Effective Altruism’ (EA), which aims to use reason and evidence to increase well-being and to decrease suffering for sentient beings such as humans and other animals. EA is a fast-growing movement closely allied with the rationality, consequentialist, evidence-based charity, and animal welfare movements. Its focus is on finding global, neglected, solvable problems and developing powerful, systemic solutions, rather than addressing the suffering of one individual at a time (as is often done in clinical psychology and medical practice).

The problem with trying to do good is that, our evolved minds are not very good at figuring out who needs our help the most, or how best to help given limited resources. We’ll explore the psychology of altruism, utilitarianism, rationality, and bias. The scope of this course will range from concrete career guidance about how to do the most good with your life in the 21st century, given your values and abilities, to how to promote the ultra-long-term sustainability and flourishing of human and post-human life at the galactic scale.

## Overview

## This is one of the first classes on Effective Altruism ever taught at an American university. (I’ve taught it twice before at UNM, in spring 2018 and spring 2019). It will be a very interdisciplinary class, drawing ideas, insights, and findings from a wide range of fields and perspectives, including insights from:

* evolutionary psychology, evolutionary biology, biological anthropology, sociology, economics, and behavior genetics.
* the rationality movement, including judgment and decision-making research, Bayesian updating, debiasing training, critical self-reflection, and randomized clinical trials
* moral philosophy, especially utilitarianism, and critiques of deontological ethics, virtue ethics, and moral signaling; the psychology utilitarian reasoning
* moral psychology research on moral foundations, the moral circle, empathy vs. rational compassion, and ethical dissorance
* research on charity effectiveness, ethical consumerism, ethical investment, and the psychology of charity, altruism, and consumer choice
* research on global poverty and public health
* research on existential risks, including nuclear war, engineered pandemics and bioweapons, autonomous lethal weapons, psychological challenges in thinking about X-risks and far-future well-being
* artificial intelligence safety, X-risks, machine ethics, and anthropomorphism
* moral and cognitive enhancement; embryo selection, biomedical interventions, applied bioethics
* ethics of autonomous humanoid robots, whole brain emulations, virtual reality, the simulation hypothesis, and the Singularity
* animal sentience, behavior, and welfare; factory farming, veganism, speciesism, the logic of the larder, wild animal suffering
* career guidance research on the most effective ways to do good through your choice of work, social network, and lifestyle
* the future of altruism; ethics for possible short-term and long-term futures

The class may be especially interesting to students who are unsure what to do with their lives and careers. We’ll spend a fair amount of discussing some options you may not have considered, that could result in a much bigger positive impact than more obvious choices.

  Readings will be a mixture of journal papers, book chapters, and popular articles; grades will be based mostly on class discussion and one term paper completed in three stages. We may have a few guest lectures from local and visiting Effective Altruism experts such as Diana Fleischman (University of Portsmouth).

**Course mechanics**

## We will meet once a week for two and a half hours. I expect you to arrive punctually, with bags unpacked, readings in front of you, and your brain ready to participate, by the start of class. There will be a 10-minute break about half way through each class.

## If you have to miss a class for any reason, please let me know by email as soon as you know you’ll be absent. Unexplained absences will reduce your grade by reducing your class participation score.

**Required book to buy:**

# *Doing Good Better: How Effective Altruism Can Help You Help Others, Do Work that Matters, and Make Smarter Choices about Giving Back* (2016) by William MacAskill.

* From UNM bookstore for $17 new or Amazon.com for about $11 new paperback, or $12 Kindle <https://amzn.to/2YVi7cX>
* This is the simplest, most accessible introduction to EA, by the Oxford professor of moral philosophy, Will MacAskill, who co-founded the EA movement. It’s about 200 pages long, and we’ll read most of it.

**Readings**

Each week there will be 3 to 5 required readings, mostly journal papers, book chapters, technical reports, or EA blogs, which should take about 3 hours on average to read or watch.

*Please do not take this course if you cannot commit about 3 hours a week to doing the readings and viewings*. The educational benefits of the course depend on you doing the readings and viewings on time, so you can contribute to class discussions. If you don’t read the assignments, you won’t learn much; if you do read them attentively, you’ll learn a lot. Some of the assignments are harder than others; some weeks require more time than other weeks. You won’t be expected to do any readings before the first class.

In the schedule of topics and readings later in this syllabus, I’ve listed both the required readings for each week, and also some optional readings that might interest some of you, or that might be useful for your term papers. I’ve also listed some suggested viewing, including a few TED-style talks and some science fiction films and ‘Black Mirror’ episodes related to the course content.

I will upload the required readings and most of the optional readings to the course’s UNM Learn site. I’ve also included web links for all readings and videos. I recommend printing out hardcopies so you can underline and take notes more easily. You can keep the printed papers in a 3-ring binder for easy reference. There will be a total of about 35 papers and other shorter pieces, plus one 200-page book, to read during the term.

I expect all of each week’s required readings to be completed well before class, so you have time to digest them, think about them, compare and contrast them, and prepare intelligent comments and questions about them. Last-minute reading will not result in good comprehension or good in-class discussion. If you see a word, term, or concept that you don’t understand when reading, don’t just gloss over it and hope for the best; instead, look it up through Google search or Wikipedia. I’ve tried to find good, short, recent, interesting readings that aren’t too technical, but there will be some jargon that you’ll need to look up.

In the course schedule on the last few pages of this syllabus, I’ve included page-counts of actual text to read (not including references in the bibliography) for each assignment. The papers will probably take 3-4 minutes per page; the main textbook may be a bit faster. Plan your study time accordingly.

**Grading.**

There will be no tests or exams in this course. Instead, your grade will depend on two kinds of work.

* **40%: class attendance, participation, and discussion points** (see below)
* **60% of grade: one term paper**, due in three stages (see below)

# Class attendance, participation, and discussion points: 40% of final grade

Most of our class time will be spent discussing the readings. Sometimes I’ll give little mini-lectures for a few minutes on particular ideas or findings that might need explaining from the readings. But for most of each class, I’ll be moderating discussions and debates among you, the students.

So, I expect regular attendance, knowledge of assigned readings, active participation and intellectual engagement, and thoughtful questions and commentaries about the readings. I will keep records of who attends each class, who asks good questions and makes insightful comments, who seem to have done the readings conscientiously, and who contributes to making this class an intellectually vibrant and rewarding experience for the other students.

**Before each class, you should write down one good discussion point about each assigned reading**. Since there are three or four assigned readings for a class, you should have at least three or four discussion points ready to go.

Each discussion point could be a thoughtful comment, question, critique, or comparison to other readings, theories, or findings. It should not just summarize the reading’s argument, but it should show that you have understood the reading, and developed your own thoughts in response. It should not just be a personal reaction or anecdote vaguely related to the reading, but it could relate the reading’s ideas to current events, controversies, or real-life issue. The best discussion points are both funny and intellectually serious.

If you haven’t understood the reading well enough to prepare a discussion point, you should be ready to say what specific theories, concepts, or findings you found most confusing, and why. If you didn’t understand something after reading it carefully, other students probably didn’t either, and we should discuss and clarify it.

You should write out these discussion points before each class, expressed clearly and concisely enough that you can read them our loud quickly, and written down clearly so you can read them. I won’t collect these discussion points, but I will call on students to read them aloud to the class. If I call on you and you haven’t done that particular reading or prepared a thoughtful comment, your participation grade will be lower for that class. If I call on you and you have a great comment that sparks a lot of discussion, your participation grade will be higher.

If you’re shy, knowing that you have good discussion points written down ahead of time will make it easier for you to speak up without feeling awkward.

The best discussion points do not just show off how clever you are, but are effective at getting other students engaged in the intellectual life of our class. You’ll learn as the term progresses what kind of comments are good discussion-sparks and which fall flat.

## Term paper in 3 stages: 60% of final grade

The term paper determines 60% of your course grade, and will be developed in three stages that will be graded separately. You can choose any topic related to the course content and course readings. The final paper should be about 3,000 words, plus references – that’s about 12 pages double-spaced. I care more about clarity, insight, research, and the flow of argument than about length per se.

The final term paper must be in standard APA (American Psychological Association) research paper format. If you’re not familiar with APA style, skim the *APA Publication Manual* for details, or see one of the online sites that summarizes its key points, such as <https://owl.english.purdue.edu/owl/resource/560/01/> . All 3 stages should be printed single-sided on standard size white paper, with 1 inch margins, double-spaced, in 12 point Arial or Times Roman font.

For graduate students, my goal is for you to produce a paper that could serve as a useful part of your master’s thesis, comprehensive examination, or dissertation – or that you could submit to a decent journal as a review or commentary piece.

To make sure that you are thinking, researching, and writing the paper on a good schedule throughout the semester, you’ll submit the three assignments about one month apart. I’ve used this system in many previous classes, and it really helps students pace their work and avoid any last-minute panic about their papers.

Stage 1 (due in Class Tues Feb 25): Provisional title, abstract, and annotated bibliography due in class (10% of final course grade).

Turn in a hardcopy document in class that day, including:

1) A provisional title, ideally about 8-20 words, that concisely explains what you think you’ll write about for the rest of the term. The title should be clear, concise, detailed, and take a stand on some issue in emotions research. A bad example title would be “The end of factory farming” – it’s too vague. A better example title would be “How much will lab-grown meat reduce demand for factory-farmed meat?” If you change your mind about your topic later, no problem, just tell me in an email later. Pick a topic that you feel passionate about – you’ll have to live with it for several months! (Of course, also include your name under the title).

2) A one-paragraph abstract, just under your title, in APA format, ideally about 150-250 words. Your title alone should give me a clear idea what you expect to write about. This abstract should go into more detail, demonstrating that you’ve already been doing some background reading about your issue, and you’re already starting to develop some of your own ideas. It should give me a sense of how you’ll structure your final term paper, how your flow of argument will go, and what kind of examples and findings you’ll use to support your points. The best abstracts have no wasted words or fluff: they get straight to the point and have real detail.

3) A provisional bibliography, after the abstract, listing about 10 to 20 sources relevant to your topic that you have actually read. Most of them should be journal papers; a few could be books (if you really had time to read them); they should not just be online resources, blogs, popular science features, etc. They should not all be from the syllabus here – you should use UNM’s online library resources and Google Scholar to find additional journal papers. In citing them, use standard APA reference format, e.g. as explained here: <http://www.library.cornell.edu/resrch/citmanage/apa> .

Also, you should explain each reference’s relevance to your topic with a brief annotation right after the APA citation. A good annotation would be “This paper critically reviews 18 recent RCTs testing the effects of deworming on economic growth.” A bad annotation would be “Reviews worms and poverty”.

After I get this initial packet from you, I will write comments and suggestions on it and return it as soon as I can.

**Stage 2 (due in class Tues March 24): Revised abstract, outline, and annotated bibliography due in class (20% of course grade).**

Before doing this assignment, you should consider very carefully the comments and suggestions that I gave on your stage 1 assignment. Students who take my feedback seriously tend to get better grades.

The assignment format here is similar: turn in a hardcopy documentin class that day, printed single-sided on standard size white paper, with 1 inch margins, double-spaced, in 12 point Arial font. This assignment should include:

1) A revised title – ideally, more concise, detailed, and exciting than before, but still about 8-20 words; include your name under the title.

2) A revised abstract that shows further research, thought, and organizational strategy, still about 150-250 words

3) An outline, around 2 pages long, that shows the planned structure of your paper. It should clearly show your flow of argument, and the specific theories, findings, and issues that you’ll consider.

The outline should include about 4 to 7 section headings that describe the overall organization of your term paper. A bad section heading would be a vague place-holder, such as “Introduction” or “Conclusion”. A good section heading would be much meatier, e.g. “Introduction: Three cognitive obstacles to consequentialist thinking about long-term X-risk”.

Under each section heading, you should have outline entries that show how you’ll flesh out your arguments. A bad entry would be “Singer’s ideas.” A good entry would be “Singer’s ideas about transhumanist bioethics”. Outline entries can also cite bibliography entries as above.

4) A revised annotated bibliography. This should include about 20 to 30 references that you have actually read, and that you plan to cite in the final paper.

After I get this revised packet from you, I will write comments and suggestions on it and return it as soon as I can. This should allow you to submit a really good final draft, and I hope it will help you improve your writing generally.

##### Stage 3 (due Tues May 5 in class): Final term paper due (30% of course grade):

This should be the culmination of three months of research, thinking, and writing about a topic that passionately interests you. It should take seriously my feedback on the stage 1 and 2 assignments. It must be in standard APA (American Psychological Association) research paper format; see the *APA Publication Manual* for details. This means double-spaced, single-sided, in 12 point Arial font, with a proper title page, abstract, references, and page numbering. It should be a well-polished document, thoroughly proofread, with very few spelling or grammatical errors.

The final term paper should include the following:

* Title: a clear, descriptive, engaging title, about 8-20 words, and your name
* Abstract: a concise, punchy abstract that interests the reader in your paper, about 150-250 words
* Introduction: Start with a bang. Pose the problem that interests you, and how you’ll approach it. Say where you stand, and why the reader should care. Be specific and clear; mix the theoretical and methodological level of discourse with real-life examples and issues; know when to be funny and when to be serious.
* Body of the paper: depending on what you’re writing about, this could include a literature review, a series of arguments, an overview of relevant ideas and research from a related area or field, a series of methodological analyses, criticism, and suggestions, or anything that advances your points. If you include literature reviews, don’t do generic overviews – review the literature with a purpose, critically, as it pertains to your topic.
* Research proposal: towards the end of your paper, sketch out a new empirical study that could resolve one of the issues you’ve raised in your paper. This could be a brief outline of a proposed experiment, an observational study, analysis of an archival dataset, randomized controlled field trial, technical AI safety proposal, or any other approach you think would be appropriate. You should explain what data would be collected, how it would be analyzed (roughly), and how the results would give insight into one of your paper’s key outstanding questions.
* Annotated Bibliography: This should include about 20 to 30 good, relevant references that you’ve actually read; only some of them should be from this class syllabus. If your bibliography includes good, relevant papers and books that I haven’t read before, I will be impressed.

**Schedule of topics and readings week by week**

## Class 1: Jan 21: Course overview and mechanics

No assigned readings before first class.

## Class 2: Jan 28: Introduction to Effective Altruism

**Required readings to complete before this class (46 pp total):**

* MacAskill Intro: worms and water pumps (p. 1-14; 14 pp)
* MacAskill Ch. 1: You are the 1 percent (p. 15-25; 11 pp)
* MacAskill Ch. 2: Q1: How many people benefit? (pp. 29-42; 14 pp)
* [Effective Altruism FAQ](https://whatiseffectivealtruism.com/) (7 pp):

Optional reading:

* Thompson, D. (2015). [The greatest good](https://www.theatlantic.com/business/archive/2015/06/what-is-the-greatest-good/395768/). *The Atlantic* magazine
* Carey, R. (2015). [*The Effective Altruism Handbook*](http://careyryan.com/files/EA_Handbook.pdf)*.* (139 pp).
* Gabriel, I. (2017). [Effective Altruism and its critics](http://onlinelibrary.wiley.com/doi/10.1111/japp.12176/abstract). *J. of Applied Philosophy, 34*(4), 457-473.
* Clough, E. (2015). [Effective Altruism’s political blind spot](http://bostonreview.net/world/emily-clough-effective-altruism-ngos). *Boston Review*.
* Explore the websites of [Effective Altruism](https://www.effectivealtruism.org/) and the [Effective Altruism Forum](http://effective-altruism.com/)

Optional viewing:

* Peter Singer (2013). [The how and why of effective altruism](https://www.ted.com/talks/peter_singer_the_why_and_how_of_effective_altruism). TED talk. (17 mins):
* Will MacAskill (2017) [opening talk](https://www.youtube.com/watch?v=HOWb3ycqFa0) at EA Global 2017 (40 mins)
* Toby Ord (2014). [How to save hundreds of lives](https://www.youtube.com/watch?v=ZGAkrpwyu1k). TEDxCambridgeUniversity. (16 mins)

## Class 3: Feb 4: Cause prioritization: Evidence-based analysis

**Required readings (62 pp total):**

* MacAskill Ch. 3: Q2: Is this the most effective thing you can do? (p. 43-54; 12 pp)
* MacAskill Ch. 4: Q3: Is this area neglected? (pp. 55-66; 12 pp)
* MacAskill Ch. 5: Q4: What would have happened otherwise? (pp. 67-78; 12 pp)
* MacAskill Ch. 6: Q5: What are the chances of success? (pp. 79-99; 21 pp)
* MacAskill Ch. 10: Which causes are most important? (pp. 179-195; 17 pp)

Optional reading

* Explore the website of [The Life You Can Save](https://www.thelifeyoucansave.org/)
* Karnofsky, H. (2015). [The lack of controversy over well-targeted aid](https://blog.givewell.org/2015/11/06/the-lack-of-controversy-over-well-targeted-aid/). *The GiveWell Blog*.
* McMahan, J. (2016). [Philosophical critiques of Effective Altruism](http://jeffersonmcmahan.com/wp-content/uploads/2012/11/Philosophical-Critiques-of-Effective-Altruism-refs-in-text.pdf). *The Philosophers’ Magazine* Issue 73, pp. 92-99. (8 pp).
* Pummer, T. (2016). [Whether and where to give](https://doi.org/10.1111/PAPA.12065). *Philosophy & Public Affairs, 44*(1), 77–95.

Optional viewing:

* Holden Karnofsky (2015). [The Open Philanthropy Project](https://www.youtube.com/watch?v=AOvUQnCMwww). EA Global talk. (25 mins).
* Julia Galef (2016). [Why you think you’re right](https://www.youtube.com/watch?v=w4RLfVxTGH4&t=2s). TED talk. (11 mins)

### Class 4: Feb 11: Utilitarian philosophy and psychology

**Required readings (32 pp total):**

* Montgomerie, I. (2000). [A utilitarian FAQ](https://www.lri.fr/~dragice/utilitarianism/faq.html). [selected excerpts] (12 pp)
* Kahane, G., Everett, J. A. C., Earp, B. D., et al. (2017). [Beyond sacrificial harm: A two-dimensional model of utilitarian psychology](http://www.jimaceverett.com/wp-content/uploads/2017/12/Kahane-Everett-et-al.-2018-Psychological-Review-2D-Model-Utilitarian-Psychology.pdf). *Psychological Review*, Epub. Just read pp. 1-8 (up to ‘Scale Development Procedure’), and pp. 25 (‘General Discussion’) through Conclusion (p. 31) (14 pp total).
* Bostrom, N. (2003). [Astronomical waste: The opportunity cost of delayed technological development](https://nickbostrom.com/astronomical/waste.html). *Utilitas, 15*(3), 308-314. (6 pp).

Optional reading

* Crisp, R. (2016). [Well-Being](https://plato.stanford.edu/entries/well-being/). *The Stanford Encyclopedia of Philosophy.*
* Ord, T. (2015). [Moral trade](http://www.amirrorclear.net/files/moral-trade.pdf). *Ethics, 126*, 118-138. (21 pp)
* Sinnott-Armstrong, W. (2015). [Consequentialism](https://plato.stanford.edu/entries/consequentialism/). *Stanford Encyclopedia of Philosophy*. (16 pp).
* Smuts, A. (2014). [To be or never to have been: Anti-natalism and a life worth living](https://www.jstor.org/stable/24478599). *Ethical Theory and Moral Practice, 17*, 711-729 (18 pp)
* Woollard, F., & Howard-Snyder, F. (2016). [Doing vs. allowing harm](https://plato.stanford.edu/entries/doing-allowing/). *The Stanford Encyclopedia of Philosophy.*

Optional viewing

* Peter Singer (2017). [Utilitarianism](https://www.youtube.com/watch?v=bRPE0IImxzs). (4 mins)
* CrashCourse Philosophy (2016). [Utilitarianism](https://www.youtube.com/watch?v=-a739VjqdSI). (10 mins)

### Class 5: Feb 18: Moral psychology, the moral circle, empathy, and dissonance

**Required readings (31 pp total):**

### Haidt, J. (2013). [Moral psychology for the twenty-first century](http://www.tandfonline.com/doi/full/10.1080/03057240.2013.817327). *J. of Moral Education, 42*(3), 281-297 (15 pp)

* Graham, J., Waytz, A., Meindl, P., et al. (2017). [Centripetal and centrifugal forces in the moral circle: Competing constraints on moral learning](https://www.sciencedirect.com/science/article/pii/S0010027716302876). *Cognition, 167*, 58-65. (6 pp).
* Bloom, P. (2016). [Empathy and its discontents](https://www.ncbi.nlm.nih.gov/pubmed/27916513). *Trends in Cognitive Sciences, 21*(1), 24-31. (7 pp)
* Barkan, R., Ayal, S., & Ariely, D. (2015). [Ethical dissonance, justifications, and moral behavior](http://portal.idc.ac.il/FacultyPublication.Publication?PublicationID=4540&FacultyUserName=cy5heWFs). *Current Opinion in Psychology, 6*, 157-161. (3 pp)

Optional reading

* Doris, J., et al. (2017). [Moral psychology: Empirical approaches](https://plato.stanford.edu/entries/moral-psych-emp/). *Stanford Encyclopedia of Philosophy*.
* Everett, J. A. C., Faber, N. S., Savulescu, J., et al. (2018). [The costs of being consequentialist: Social perceptions of those who hard and help for the greater good](https://psyarxiv.com/a2kx6/). J. Exp. Social Psych, 79, 200-216. (17 pp)
* Miller, G. F. (2007). [Sexual selection for moral virtues](https://www.ncbi.nlm.nih.gov/pubmed/17583267). *Quarterly Review of Biology, 82*(2), 97-125. (21 pp)

Optional viewing

* Jonathan Haidt (2013). [Why liberals and conservatives can’t get along](https://www.youtube.com/watch?v=qN42ZLwNFBY). Wharton Business School talk. (27 mins).
* Paul Bloom (2016). [Against empathy: The case for rational compassion](https://www.youtube.com/watch?v=yhCGmDJQRpc). Carnegie Council talk. (55 mins)
* Dan Ariely (2012). [Our buggy moral code](https://www.youtube.com/watch?v=onLPDegxXx8). TED talk. (16 mins).
* Lucius Caviola (2016). [The psychology of Effective Altruism](https://www.youtube.com/watch?v=0MKrN2Iyfug). Humboldt University talk (56 mins).

**Class 6: Feb 25: Effective charities, ethical consumerism, and virtue-signaling**

##### *\*\*Term paper stage 1 due in class: Provisional title, abstract, and bibliography*

**Required readings (55 pp total):**

* MacAskill Ch. 7: What charities make the most difference? (pp. 103-127; 24 pp)
* MacAskill Ch. 8: How can consumers make the most differences? (pp. 128-146; 19 pp)
* Miller, G. F. (2013). [Twenty-seven thoughts about multiple selves, sustainable consumption, and human evolution](https://psych.unm.edu/people/faculty/profile/miller-geoffrey-papers/2013-sustainable-consumerism.pdf) (pp. 27-35). In H. C. M. van Trijp (Ed.), *Encouraging sustainable behavior: Psychology and the Environment*.  Psychology Press.  (6 pp)
* Miller, G. F. (2000). [Moral vision](https://www.primalpoly.com/s/2000-moral-vision.doc): Effective altruism and ethical investment through augmented reality. (6 pp)

Optional reading

##### Caviola, L., Faulmuller, N., Everett, J. A. C., et al. (2014). [The evaluability bias in charitable giving: Saving administrative costs or saving lives](https://www.academia.edu/7798644/The_evaluability_bias_in_charitable_giving_Saving_administration_costs_or_saving_lives)? *Judgment and Decision Making, 9*(4), 303-315. (10 pp).

##### Griskevicius, V., Cantu, S. M., & van Vugt, M. (2012). [The evolutionary bases for sustainable behavior: Implications for marketing, policy, and social entrepreneurship](https://pdfs.semanticscholar.org/0b55/6a228bf31d54df996755e223ebc95c11e3de.pdf). *J. of Public Policy & Marketing, 31*(1), 115-128. (11 pp)

* Miller, G. F. (2012). [Sex, mutations, and marketing](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3463973/). *EMBO Reports, 13*(10), 880-884. (5 pp)
* Explore effective charity evaluation at the websites of [GiveWell](https://www.givewell.org/), [Open Philanthropy Project](https://www.openphilanthropy.org/), [Charity Science Foundation](http://www.charityscience.com/),

Optional viewing

* The ‘Black Mirror’ episode ‘[Nosedive](https://en.wikipedia.org/wiki/Nosedive)’ concerns virtue-signaling

**Class 7: March 3: Global poverty and health**

**Required readings (46 pp total):**

* Wenar, L. (2010). [Poverty is no pond: Challenges for the affluent](http://files.givewell.org/files/DWDA%202009/Interventions/Wenar%202010.pdf). In P. Illingworth et al. (Eds), *Giving Well: The Ethics of Philanthropy*, pp. 104-132. Oxford U. Press. (28 pp).
* Ord, T. (2013). [The moral imperative toward cost-effectiveness in global health](https://www.cgdev.org/publication/moral-imperative-toward-cost-effectiveness-global-health). Center for Global Development technical report. Read pp. 1-7. (7 pp)
* Mitra, A. K., & Mawson, A. R. (2017). [Neglected tropical diseases: Epidemiology and global burden](http://www.mdpi.com/2414-6366/2/3/36/pdf). *Tropical Medicine and Infectious Disease, 2,* 36. (11 pp).

Optional reading:

* Explore direct cash transfer programs at [GiveDirectly](https://www.givedirectly.org/) , efficient remittances at [Wave](https://www.wave.com/) , and public health outreach at [Development Media International](http://www.developmentmedia.net/)
* Arrhenius, G., Ryberg, J., & Tännsjö, T. (2017). [The repugnant conclusion](https://plato.stanford.edu/entries/repugnant-conclusion/). *Stanford Encyclopedia of Philosophy.*
* Ashford, E. (2009). [Obligations of justice and beneficence towards the severely poor](http://www.oxfordscholarship.com/view/10.1093/acprof%3Aoso/9780199739073.001.0001/acprof-9780199739073-chapter-3). In T. Pogge et al. (Eds.), *The Ethics of Philanthropy.* Oxford U. Press.
* Barry, C., & Øverland, G. (2016). [*Responding to global poverty: Harm, responsibility, and agency*](https://www.cambridge.org/core/books/responding-to-global-poverty/67CEF5165F08653587497D5F9A4118E9)*.* Cambridge University Press.
* Saavedra, J. E. (2016). [The effects of conditional cash transfer programs on poverty reduction, human capital accumulation and wellbeing](http://www.un.org/esa/socdev/egms/docs/2016/Poverty-SDGs/JuanSaavedra-paper.pdf). United Nations Division for Social Policy and Development report. (8 pp)
* Wiblin, R. (2015). [Effective altruists love systemic change](https://80000hours.org/2015/07/effective-altruists-love-systemic-change/). *80,000 Hours Blog.*

Optional viewing:

* Eva Vivalt (2015). [Global poverty](https://www.youtube.com/watch?v=MY_oC-Dxb4k). EA Global 2015. (20 mins).
* Grace Hollister (2015). [Let’s deworm the world](https://www.youtube.com/watch?v=MsjUKbOGn8U). EA Global Melbourne. (30 mins).
* Kevin Esvelt (2016). [Beyond malaria: Gene drives, global health, and how we do science](https://www.youtube.com/watch?v=T1wK0rSn3CM). EA Global 2016 (35 mins)

**Class 8: March 10: Existential risks: Psychology and practicalities**

**Required readings (65 pp total):**

* Global Priorities Project (2017). An introduction to existential risks (section 1, pp. 6-12) in [*Existential Risk: Diplomacy and governance*](https://www.fhi.ox.ac.uk/wp-content/uploads/Existential-Risks-2017-01-23.pdf) (7 pp)
* Global Challenges Foundation (2017). Sections on specific existential risks (pp. 12-44) in [*Global Catastrophic Risks 2017*](https://api.globalchallenges.org/static/files/Global%20Catastrophic%20Risks%202017%20BW.pdf) (33 pp)
* Yudkowsky, E. (2008). [Cognitive biases potentially affecting judgment of global risks](https://intelligence.org/files/CognitiveBiases.pdf). In N. Bostrom & Cirkovic, M. M. (Eds.), *Global catastrophic risks*, pp. 91-119. Oxford U. Press. (25 pp).

Optional readings

* Explore research on X-risk at websites of Future of Humanity Institute ([FHI](https://www.fhi.ox.ac.uk/)), Cambridge Centre for the Study of Existential Risk ([CSER](https://www.cser.ac.uk/)), Foundational Research Institute ([FRI](https://foundational-research.org/)), Future of Life Institute ([FLI](https://futureoflife.org/)), Global Priorities Project ([GPP](http://globalprioritiesproject.org/))
* Beckstead, N. (2013). [*On the overwhelming importance of shaping the far future*](https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbnxuYmVja3N0ZWFkfGd4OjExNDBjZTcwNjMxMzRmZGE)*.* (PhD Dissertation, Rutgers University).
* Beckstead, N. (2015). [How much could refuges help us recover from a global catastrophe?](https://www.sciencedirect.com/science/article/pii/S0016328714001888) *Futures, 72*, 36-44. (8 pp)
* Bostrom, N. (2013). [Existential risk prevention as global priority](http://www.existential-risk.org/concept.pdf). *Global Policy, 4*(1), 15-31. (12 pp)
* Meyer, L. (2015). [Intergenerational justice](https://plato.stanford.edu/entries/justice-intergenerational/). The Stanford Encyclopedia of Philosophy
* Millett, P., & Snyder-Beattie, A. (2017). [Existential risk and cost-effective biosecurity](https://www.ncbi.nlm.nih.gov/pubmed/28806130). *Health Security, 15*(4), 373-383. (9 pp).
* Mulgan, T. (2015). [Utilitarianism for a broken world](https://www.cambridge.org/core/journals/utilitas/article/utilitarianism-for-a-broken-world/5747D72DC6888D3F22C21969511890D4). *Utilitas*, 27, 92-114.

Optional viewing:

* Peter Singer (2014). [Extinction risk and effective altruism](https://www.youtube.com/watch?v=uh8nFrp9FFI). (6 mins).
* Nick Bostrom (2013). [The end of humanity](https://www.youtube.com/watch?v=P0Nf3TcMiHo). TEDxOxford talk. (16 mins).
* Max Tegmark (2017). [Effective altruism, existential risk, and existential hope](https://www.youtube.com/watch?v=2f1lmNqbgrk). EA Global talk. (35 mins).
* Niel Bowerman (2015). [Astronomical stakes](https://www.youtube.com/watch?v=p0JPpRswZKY). EA Global Melbourne. (34 mins).

**(No class March 17: UNM Spring Break)**

**Class 9: March 24: Artificial Intelligence as an X-risk: Psychology and ethics**

##### *\*\*Term paper stage 2 due in class: Revised abstract, outline, and bibliography*

**Required readings (34 pp total):**

* Futurism Staff (2017). [Interview with James Barrat](https://futurism.media/interview-with-our-final-invention-author-james-barrat) (4 pp)
* Global Challenges Foundation (2017). Sections on Artificial Intelligence (pp. 45-49 and 60-61) in [*Global Catastrophic Risks 2017*](https://api.globalchallenges.org/static/files/Global%20Catastrophic%20Risks%202017%20BW.pdf) (7 pp)
* Bostrom, N., & Yudkowsky, E. (2014). [The ethics of artificial intelligence](https://nickbostrom.com/ethics/artificial-intelligence.pdf). In K. Frankish & W. M. Ransay (Eds.), *The Cambridge Handbook of Artificial Intelligence* (pp. 316-334). NY: Cambridge U. Press. (17 pp)
* Proudfoot, D. (2011). [Anthropomorphism and AI: Turing’s much misunderstood imitation game](https://www.sciencedirect.com/science/article/pii/S000437021100018X). *Artificial Intelligence, 175*, 950-957. (6 pp).

Optional readings:

* [Asilomar AI Principles](https://futureoflife.org/ai-principles/) (2017) (2 pp)
* Explore AI ethics research at the websites of Machine Intelligence Research Institute ([MIRI](https://intelligence.org/)), Future of Humanity Institute ([FHI](https://www.fhi.ox.ac.uk/)), Cambridge Centre for the Study of Existential Risk ([CSER](https://www.cser.ac.uk/)), [OpenAI](https://openai.com/), [DeepMind Ethics & Society](https://deepmind.com/applied/deepmind-ethics-society/)
* Amodei, D., Olah, C., Steinhardt, J., et al. (2016). Concrete problems in AI safety. [arXiv:1606.06565v2](https://arxiv.org/abs/1606.06565)
* Bostrom, N. (2012). [The superintelligent will: Motivation and instrumental rationality in advanced artificial agents](https://nickbostrom.com/superintelligentwill.pdf). *Minds & Machines, 22*, 71-85.
* Russell, S., Dewey, D., & Tegmark, M. (2015). [Research priorities for robust and beneficial artificial intelligence](https://futureoflife.org/data/documents/research_priorities.pdf). *AI Magazine 36*(4). (7 pp)
* Roman V. Yampolskiy, R. I. (2014). [Artificial intelligence safety engineering: Why machine ethics is a wrong approach](http://cecs.louisville.edu/ry/AIsafety.pdf). In V.C. Müller (Ed.): *Philosophy and Theory of Artificial Intelligence, SAPERE 5*, pp. 389–396. (6 pp).

Optional viewing:

* Max Tegmark (2017). [Why superintelligent AI could be the last human invention](https://www.youtube.com/watch?v=m4qdQl6bjUU). The Big Think. (5 mins).
* Sam Harris (2016) [Can we build AI without losing control over it?](https://www.youtube.com/watch?v=8nt3edWLgIg) TED talk (14 mins)
* Nick Bostrom (2015). [What happens when our machines get smarter than we are?](https://www.youtube.com/watch?v=MnT1xgZgkpk) TED talk. (16 mins)
* Stuart Russell (2017). [Three principles for creating safer AI](https://www.youtube.com/watch?v=EBK-a94IFHY). TED talk. (18 mins).
* Best documentaries about AI: [*AlphaGo*](http://www.imdb.com/title/tt6700846/) (2017)
* Best science fiction movies about AI: [*Ex Machina*](http://www.imdb.com/title/tt0470752/) (2014), [*Her*](http://www.imdb.com/title/tt1798709/) (2013).

**Class 10: March 31: Moral and cognitive enhancement**

**Required readings (26 pp total)**

* Giubilini, A., & Sanyal, S. (2015). [The ethics of human enhancement](http://onlinelibrary.wiley.com/doi/10.1111/phc3.12208/abstract). *Philosophy Compass, 10/4*, 233-243. (8 pp)
* Shulman, C., & Bostrom, N. (2014). [Embryo selection for cognitive enhancement: Curiosity or game-changer?](https://nickbostrom.com/papers/embryo.pdf) *Global Policy, 5*(1), 85-92. (6 pp)
* Miller, G. F. (2013). [Chinese eugenics](https://www.edge.org/response-detail/23838). EDGE.org. (2 pp)
* Persson, I., & Savulescu, J. (2017). [Moral hard-wiring and moral enhancement](https://www.ncbi.nlm.nih.gov/pubmed/28300281). *Bioethics, 31*(4), 286-295. (10 pp)

Optional reading:

* Bostrom, N., & Ord, T. (2006). [The reversal test: Eliminating the status quo bias in applied ethics](https://nickbostrom.com/ethics/statusquo.pdf). *Ethics, 116*, 656-679. (24 pp)
* Harris, J., & Savulescu, J. (2015). [A debate about moral enhancement](https://www.cambridge.org/core/services/aop-cambridge-core/content/view/S0963180114000279). *Cambridge Quarterly of Healthcare Ethics, 24*(1), 8-22. (14 pp).
* Earp, B. D. (2018). [Psychedelic moral enhancement](https://www.academia.edu/33771413/Psychedelic_moral_enhancement). In M. Hauskeller & L. Coyne (Eds.), *Moral Enhancement: Critical Perspectives*, Vol. 83, 415-439 Cambridge University Press. (24 pp)

Optional viewing:

* Julian Savulescu (2013). [Pills that improve morality](https://www.youtube.com/watch?v=DhtIFTrJQJ4). TEDxBarcelona talk. (15 mins)
* John Harris (2017). [Should we edit the genome?](https://www.youtube.com/watch?v=kIMOpu9uZbk) Even for enhancement? Practical Ethics Channel. (17 mins)
* ‘Black Mirror’ episodes about enhancement technologies: ‘[The entire history of you](https://en.wikipedia.org/wiki/The_Entire_History_of_You)’, ‘[Men against fire’](https://en.wikipedia.org/wiki/Men_Against_Fire), ‘[Arkangel](https://en.wikipedia.org/wiki/Arkangel_%28Black_Mirror%29)’

### Class 11: April 7: Ethics of robots, brain emulations, and virtual reality

**Required readings (36 pp total):**

##### Danaher, J., Earp, B. D., & Sandberg, A. (in press). [Should we campaign against sex robots?](https://www.academia.edu/25063138/Should_we_campaign_against_sex_robots) In J. Danaher & N. McArthur (Eds.) *Robot sex: Social and ethical implications.* MIT Press. (25 pp)

* Hanson, R. (2014) [What will it be like to be an Emulation?](http://mason.gmu.edu/~rhanson/IntelligenceUnbound.pdf) In R. Blackford & D. Broderick (Eds.), *Intelligence unbound: The future of uploaded and machine minds*, pp. 298-309. Wiley. (9 pp).
* Miller, G. F. (2007). [Runaway consumerism explains the Fermi paradox](https://www.edge.org/response-detail/11475). In J. Brockman (Ed.), *What is your dangerous idea?* (pp. 240-243). Harper Perennial. (2 pp)

Optional reading

* Mulgan, T. (2014). [Moral philosophy, superintelligence, and the singularity](https://www.academia.edu/16674139/Moral_Philosophy_Superintelligence_and_the_Singularity). Unpublished talk. (17 pp)
* Sotala, K., & Gloor, L. (2017). [Superintelligence as a cause or cure for risks of astronomical suffering](http://www.informatica.si/index.php/informatica/article/view/1877). *Informatica, 41*, 501-515. (11 pp)
* Tomasik, B. (2016). [How the simulation argument dampens future fanaticism](https://foundational-research.org/how-the-simulation-argument-dampens-future-fanaticism). Foundational Research Institute technical report. (26 pp)

Optional viewing:

* Robin Hanson (2017). [What would happen if we upload our brains to computers?](https://www.youtube.com/watch?v=Urk3xn7l3AM) TED talk (12 mins).
* Anders Sandberg (2016). [Whole brain emulation](https://www.youtube.com/watch?v=A0_SE9xwqUg). (11 mins)
* Several ‘Black Mirror’ episodes concern life as an Em (an uploaded mind): ‘[San Junipero](https://en.wikipedia.org/wiki/San_Junipero)’, ‘[USS Callister](https://en.wikipedia.org/wiki/USS_Callister)’, ‘[Hang the DJ](https://en.wikipedia.org/wiki/Hang_the_DJ_%28Black_Mirror%29)’, ‘[White christmas](https://en.wikipedia.org/wiki/White_Christmas_%28Black_Mirror%29)’,
* Best movies about Ems and post-humans: [The Matrix](http://www.imdb.com/title/tt0133093/) (1999), [Blade Runner 2049](http://www.imdb.com/title/tt1856101/) (2017).

### Class 12: April 14: Animal sentience and welfare I

**Required readings (29 pp total):**

* Singer, P. (1979). Equality for animals? Chapter 3 in [*Practical Ethics*](https://en.wikipedia.org/wiki/Practical_Ethics). Cambridge U. Press. (13 pp)
* Humane Society of the United States (2009). [*The welfare of animals in the meat, egg, and dairy industries*](http://www.humanesociety.org/assets/pdfs/farm/welfare_overview.pdf). Read pp. 1-5 (5 pp)
* Braithwaite, V. A., & Ebbesson, L. O. E. (2014). [Pain and stress responses in farmed fish](https://pdfs.semanticscholar.org/82d5/49924246788e525684ef5f119c318b9219a4.pdf). *Rev. sci. tech. Off. int. Epiz., 33*(1), 245-253. (5 pp)
* Fleischman, D. (2013). [The ethical case for eating oysters and mussels](https://sentientist.org/2013/05/20/the-ethical-case-for-eating-oysters-and-mussels/), Parts 1 and 2. Sentientist blog. (6 pp)

Optional readings

* Explore the websites of [Mercy for Animals](https://www.mercyforanimals.org/), [Vegan Outreach](https://veganoutreach.org/) , [Animal Charity Evaluators](https://animalcharityevaluators.org/) , [Sentience Institute](https://www.sentienceinstitute.org/)
* Fisher, A. (2017). [Theory-neutral arguments for “effective animal advocacy](http://dx.doi.org/10.7710/1526-0569.1578)”. *Essays in Philosophy, 18*(1), 3.
* Kasperbauer, T. J. (2017). [Mentalizing animals: Implications for moral psychology and animal ethics](https://link.springer.com/article/10.1007/s11098-016-0692-5). *Philosophical Studies, 174*, 465-484. (18 pp)

Optional viewing:

* Kelly Witwicki (2016). [Effective animal activism](https://www.youtube.com/watch?v=6MzgPtyJOQk). EA Global X Berlin talk. (25 mins).
* Natalie Cargill (2016). [Political and legal activism for all sentient beings](https://www.youtube.com/watch?v=RkQtYj_az10&t=264s). EA Global X Berlin talk. (27 mins).
* Best documentaries about animal welfare issues: [Earthlings](http://www.imdb.com/title/tt0358456/) (2005), [Food, Inc](http://www.imdb.com/title/tt1286537/). (2008), [Vegucated](http://www.imdb.com/title/tt1814930/) (2011).

### Class 13: April 21: Animal sentience and welfare II

 Required readings (41 pp)

* Caviola, L., Everett, J. A. C., & Faber, N. S. (2019). [The moral standing of animals: towards a psychology of speciesism](http://www.jimaceverett.com/publications/the-moral-standing-of-animals-towards-a-psychology-of-speciesism/). *J. of Personality and Social Psychology*, 116,1011-1029 (27 pp). (Just read pp 1-9 and 25-27 (12 pp); skip the studies)
* Matheny, G., & Chain, K. M. A. (2005). [Human diets and animal welfare: The illogic of the larder](https://link.springer.com/article/10.1007/s10806-005-1805-x). *J. of Agricultural and Environmental Ethics, 18*, 579-594. (12 pp)
* Tomasik, B. (2015). [The importance of wild-animal suffering](https://foundational-research.org/files/the-importance-of-wild-animal-suffering.pdf). *Relations: Beyond Anthropocentrism, 3*(2). (17 pp)

### Class 14: April 28: Career choices and life strategies

**Required readings (32 pp total plus online exercises):**

* MacAskill Ch. 9: Which careers make the most difference? (pp. 147-178; 32 pp)
* Do the quick 6-question [career quiz](https://80000hours.org/career-quiz/#/) on the 80,000 hours website: This takes about 5 minutes.
* Do the career [planning exercise](https://80000hours.org/career-decision/) on the 80,000 hours website:. This takes about 30 minutes.

Optional reading

* Explore the [80,000 hours](https://80000hours.org/career-decision/) website, especially their Career Guide.
* Paul, L. A. (2014). [*Transformative experience*](https://www.amazon.com/Transformative-Experience-L-Paul/dp/0198777310). Oxford University Press.
* Tomasik, B. (2017). [Education matters for altruism](file:///F%3A%5CUsers%5Cgfmiller%5CDocuments%5Cteaching%5Cspring%202018%20effective%20altruism%5C%E2%80%A2%09https%3A%5Cfoundational-research.org%5Ceducation-matters-for-altruism%5C). Blog for Foundational Research Institute.

Optional viewing

* Benjamin Todd (2015). [To find work you love, don’t follow your passion](https://www.youtube.com/watch?v=MKlx1DLa9EA). TEDxYouth@Tallinn. (15 mins)
* William MacAskill (2015). [Want to make a difference? Don’t work for a charity](https://www.youtube.com/watch?v=hU5O4PhrOik&t=2s). TEDxCambridgeUniversity. (17 mins)

**Class 15: May 5: Last day of class: The future of altruism**

##### *\*\*Term paper stage 3 due in class: Final term paper*

**Required readings (30 pp total):**

* MacAskill Conclusion: What should you do right now? (pp. 196-199 (3 pp)
* MacAskill Appendix: The five key questions of effective altruism (pp. 201-204 (4 pp)
* Mulgan, T. (2014). [Ethics for Possible Futures](http://onlinelibrary.wiley.com/doi/10.1111/j.1467-9264.2014.00364.x/abstract). *Proceedings of the Aristotelian Society*, 114, 57-73. (15 pp)
* Todd, B. (2016). [The value of coordination](https://80000hours.org/2016/02/the-value-of-coordination/). (8 pp)

**Optional reading**

* Explore the websites of [Giving What We Can](https://www.givingwhatwecan.org/) , [EA Global](https://www.eaglobal.org/) ,

# (UNM final exams are May 11-15, but there’s no final exam in this class)