Law x Effective Altruism Intro Syllabus

McGill's Law x Effective Altruism Group   Fall 2022

The Law x EA reading group aims to bring law students into the study of the **world's most pressing problems** and what we can do about them. Ideas around these issues will be discussed over eight weeks in an hour and a half sessions from the following readings.

Some of these readings (the longer ones in particular) are not expected to be read entirely. The objective is to **get a general idea of the subject** to discuss it as a group. So, if it is too much material for one week, you can choose the readings that interest you most. We suggest reading the readings in the order in which they are presented.

The reading plan is also subject to change according to the interests of the group participants, so feel **free to suggest readings or topics you would like to discuss.**

You can send anonymous comments and suggestions about the reading group and its content at any time in the following form: https://airtable.com/shrhpHmDRsgtlvZVR

Contact: law@effective-altruism-mcgill.org, Discord

Table of Contents

Acknowledgments

This curriculum is based on the one prepared by Matt, John, Ben, Trevor, Christina from 80K, the Center for Effective Altruism, and the Legal Priorities Project, and modified with Matt's permission. It is loosely based on the Legal Priorities Research Syllabus, the Legal Priorities Research Reading Group. It also pulls from those curriculums found in CEA's Intro Program - Organizer Pack.

What the reading group involves

We'll meet weekly in groups of 5-8 for 75 minutes to discuss the week's reading materials. Each week's materials include (i) core materials, (ii) optional reading material, (iii) discussion materials (criticisms and questions) and, optionally, (iv) some exercises. Everyone is expected to understand the readings' essence by the time we meet; no one is expected to read everything word for word or to take copious notes. If you don't do the readings, come anyway: most of our learning comes from our discussion and interactions, so you will still profit from being around.

- The core materials should take about .5-1 hour to skim through.

- The questions, criticisms and exercises are meant to help you make the key concepts from the readings more concrete. They can also help you consider further study and possible projects in those areas.

- All the 'Optional Reading's are – well – optional. They are meant to let you explore those themes that especially interest you in more depth and breadth.

How we hope you'll approach the reading group[1]

We hope you take ideas seriously.

We often have conversations just for fun: we bat around interesting thoughts and say smart things, then go back to doing whatever we were already doing in our lives. This is a fine thing to do — but at least sometimes when we encounter new ideas, we should be asking ourselves questions like:

- "What is this idea about and what argument is the author making?"

- "How could I tell if this idea was true?"

- "If it is true, does it mean I should be doing something differently in my life? What other assumptions or opinions should I reconsider?"

And, zooming out:

- "Where are my blind spots?"

- "What assumptions do I have? Are these assumptions true?"

- "Which important questions should I be thinking about that I'm not?"

---

[1] Inspired by Julia Galef's [Update Project](#)

Taking ideas seriously means wanting to make our worldviews as full and accurate as possible since we see that having carefully considered beliefs allows us to make better decisions about things we care about.

We hope you find disagreements interesting.

When thoughtful people with the same information reach very different conclusions, we should ask why. Often, we tend to be incurious about this simply because disagreements are so common that we don't often examine them.

If, for example, a medical community were divided on which treatment is more effective at curing a given disease, they would examine their facts, assumptions, and thought processes because the right answer matters. We can do the same, even (and especially) in areas where there's no right or wrong answer. Examining disagreements can uncover hidden beliefs and foster greater understanding, empathy, and collaboration.

We hope you present strong opinions that you hold weakly.

Often, people avoid voicing opinions and say things like "I'm not an expert" or "It's hard to know for sure." For this reading group, please be bold and venture guesses. A clearly expressed idea makes it easy for others or measurable facts to assess it. In the long run, we hope that you will become more confident and curious thinkers, so please be okay with making mistakes in the short run.

We hope you brainstorm for further action/development.

Law has not been a well-explored discipline within the EA community; while new and actionable ideas are urgently needed across all EA-oriented disciplines, it is especially so

here. Many highly-engaged EAs pursue legal education out of a desire to influence policy and work on the world's most pressing problems; almost all of them struggle to form a comprehensive plan to leverage their skills to do the most good. Bringing thoughtful, law-adjacent minds together to develop new ideas and identify needs is a principal aim of this group.
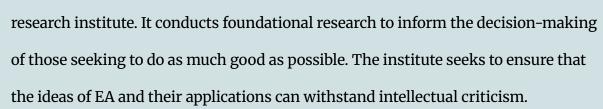
Week 1: Introductions and Presentation

Welcome! We're so glad you're here. In Week 1, we generally explore what Effective Altruism is all about through a presentation from your hosts. That being said, this is also meant to be a time for insightful introductions, curious questions, and good company.

Organization Spotlight

Global Priorities Institute

The Global Priorities Institute (GPI) is a multidisciplinary research institute. It conducts foundational research to inform the decision-making of those seeking to do as much good as possible. The institute seeks to ensure that the ideas of EA and their applications can withstand intellectual criticism.

GPI's research areas include:

- Assessing the idea of longtermism and its applications
- Investigating value questions of how people should act when they are uncertain about what will happen and what is right
- Examining the relationships between economic growth and well-being

Discussion Questions

1. How did you discover EA/this reading group? Did EA ideas influence your decision to study law?

2. On which causes do you think people with legal backgrounds can make the most positive impact?

Optional Materials

- ☐ Introduction to Effective Altruism (15 mins)
- ☐ Doing Good Better – Introduction–Chapter 1 (30 mins)
- ☐ [Benjamin Todd on the Core of Effective Altruism](#) (podcast – 84 mins)
- ☐ [Introduction to EA | Ajeya Cotra | EAGxBerkeley 2016](#) (Video – 30 mins)
- ☐ [What are the most important moral problems of our time? | Will MacAskill](#)(video – 12 mins)

Week 2: Differences in Impact

In Week 2 we continue to explore the core principles of Effective Altruism. We focus on giving you tools to quantify and evaluate how much good an intervention can achieve; introduce expected value reasoning; and investigate differences in expected cost-effectiveness between interventions.

Organization Spotlight

[GiveWell](#)

GiveWell searches for the charities that save or improve lives *the most* per dollar. They recommend a small number of charities they believe do an incredible amount of good. Unlike charity evaluators that focus solely on financials, assessing administrative or fundraising costs, they conduct in-depth research to determine how much good a given program accomplishes (in terms of lives saved, lives improved, etc.) per dollar spent.

Rather than try to rate as many charities as possible, they focus on the few charities that stand out most ([by their criteria](#)) to find and confidently recommend high-impact giving opportunities. See their [top charities here.](#)

They believe that there is exceptionally strong evidence for their top charities and that donations can save a life for every $3,000–$5,000 donated.

Core Materials

- ☐ *[Doing Good Better](#)* – Chapters 4 – 6 (45 mins)

- ☐ [One World Now Excerpt](#) – Pages 174–178, ~7 min

Exercise (30–60 mins)

Your challenge this week will be to practice generating quantitative estimates and comparing outcomes. Through this exercise, you'll attempt to estimate how much good you might be able to achieve by donating to effective charities. (Though of course this is only one of the options for having a positive impact!).

Part 1 – Estimate your likely total future income (5-10 mins)

In this part of the exercise, we'd like you to estimate what your total future income will be during your life. This is obviously quite a personal question, so this estimate is just for you, and we won't be explicitly discussing answers to this part in-session. If you'd feel

more comfortable, feel free to just estimate what an average graduate from your university will earn (for McGill Law, it's 74,000$ per year[2]).

In making this estimate, it's difficult to know what this will be and impossible to know what the future holds for you. But we think you still might be able to make educated estimates based on factors such as what the typical graduate earns, your likely career paths that you're considering now, and sense checking the answer.

Feel free to do any research that you would like to make your estimate. If you're feeling stuck, we have some tips below.

> Write your answer here (In your own copy of the document)
>
> *We're trying to put a number on the total income you'll earn over the course of your life*

If you're feeling stuck, here are some tips:
- If you have no clue about what career you'll likely pursue, try running the estimate for a typical graduate at your university instead of for yourself
- Sometimes life can throw you curveballs and mess up your plans. Try making worst case scenario, most likely scenario, and best case scenario estimates if you're feeling uncertain about what the future holds
- Break the question down eg. you might find it useful to start by estimating how many years you'll work before retirement

---

[2] According to Emolument, https://www.emolument.com/salary-reports/universities/mcgill-university/14305

- Don't worry about this question too much and try not to spend more than 10 to 15 minutes on it. It's okay to just go with a very rough and inaccurate guess.

You might want to plug in this value into Giving What We Can's [How Rich Am I?](#) calculator to see how this average annual income compares to the rest of the world.

Part 2 – What could you achieve with your income? (10-20 mins)

For the second part of the exercise we'll try and work out what you could achieve by donating some of the money you'll earn in the future.

[GiveWell](#) is an effective altruism-inspired organization which attempts to identify outstanding donation opportunities in global health and development. Using [their reports on their top charities](#) and your earlier estimate of your future income, try and work out what you could achieve if you donated 10% of your lifetime income to one of these charities[3].

If you're short on time, [here's a cheat sheet with information about three top GiveWell charities](#). If you'd like to explore further, check out GiveWell's [cost effectiveness models](#).

Complete this exercise for three GiveWell charities.

> Write your answer here (In your own copy of the document)
>
>
> *e.g.*

---

[3] 10% is the figure of the Giving What We Can Pledge, a pledge that many involved in the Effective Altruism community have taken.

Malaria Consortium: X months of malaria prevention for one person, with an estimate of N deaths averted

GiveDirectly: $X transferred to recipients

Part 2b (10 mins)

In the last section, you ended up with a few different options, (e.g. saving the lives of 200 5-year olds, doubling the income of 1000 people earning $1/day, or a 10% chance of enabling 1000 kids to attend school). Now imagine you get to donate to one of these charities.

There's a difficult judgement to be made now: since you have to pick, which charity would you donate to to do the most good?

Write your answer here (In your own copy of the document)

Which charity do you pick to donate to? Why?

Write your answer here (In your own copy of the document)

What factors (e.g. strength of evidence, predicted outcome, etc.) did you consider when picking your charity? Why?

Optional (5-10 mins)

What are other decisions in your life that you might consider generating quantitative estimates and comparing outcomes for?

(Optional) Jot your thoughts down here

Discussion Questions

1. Today's exercise

1. How did everyone find the exercise?
2. If you feel comfortable sharing, does your estimate of your total future income seem surprisingly high or low?
3. If you feel comfortable sharing, is what you could achieve with 10% of your future income more or less than you would have expected?

4. Which charity did you pick to donate to, and why did you choose it? Why did the outcomes of donating to that charity seem more valuable than the outcomes of the other charities? Did you find it hard to choose between different outcomes?

2. Benefits of cost-effectiveness estimates

5. How can we go about comparing different interventions/cause areas? Can quantitative estimates of impact be useful even if they're imprecise?
    a. How useful are QALYs?
    b. How useful is considering importance, tractability, and neglectedness?
    c. How useful is expected value (EV)?

3. Limitations of cost-effectiveness estimates

6. What types of outcomes are particularly hard to measure (or even impossible)? How should we treat such outcomes?
7. What kind of problems can we run into when we try quantifying cost-effectiveness? What features are not captured by such estimates? Are there important features of an intervention that are not captured by cost-effectiveness estimates? Which?

4. Application: Donation decisions

8. If an intervention is not backed by strong evidence, could there still be reasons to pursue it?

Optional Materials

Criticisms Against EA

- [ ] [Criticisms of EA](#) from here: [The Lessons of Effective Altruism | Ethics & International Affairs](#) (40 mins)

- [ ] [Growth and the case against randomista development](#) (60 mins - if you're short on time, read Sections 1-3)

Effective Giving

- [ ] [Our Criteria for Top Charities - GiveWell](#) and [Process for Identifying Top Charities - GiveWell](#) (20 mins)

- [ ] [Hits-based Giving - Open Philanthropy](#) (45 mins)

Evaluating and Thrust for Effective Giving

- [ ] [Expected Value](#) (2 mins)

- [ ] [Scope Insensitivity](#) (3 mins)

- [ ] [The Moral Imperative toward Cost-Effectiveness in Global Health - Centre for Global Development](#) (20 mins)

Week 3: Longtermism

In Week 2, we discussed attempting to quantify the impact of altruistic interventions. However, most cost-effectiveness analyses can only take into account the short-run effects of the interventions and struggle to take into account long-run knock-on effects and side effects. This criticism has been made forcefully against early effective altruist attempts to evaluate interventions based on cost-effectiveness.

This week we'll explore a different approach to finding high-impact interventions - 'longtermism' - which attempts to find interventions that beneficially influence the long-run course of humanity.

Organization Spotlight

[All-Party Parliamentary Group for Future Generations](#)

The All-Party Parliamentary Group (APPG) for Future Generations is a UK parliamentary group working to create cross-party dialogue on combating

short-termism and identifying ways to internalize concern for future generations into today's policy making.

They believe that political short-termism can cause topics with widespread consequences — like climate change, public health trends and catastrophic and existential risks — to be neglected from the political agenda in favour of urgent matters.

You can see their research aimed at informing Parliamentarians on catastrophic risks and potential policy options [here](). You can see their events bringing together policy, academic, and industry communities [here]().

Core Materials

- ☐ [80,000 Hours - Future Generations and Their Significance]() (20 mins)

- ☐ [Holden Karnofsky, This Can't Go On]() (15 minutes)

Exercise (10 min, please complete this before your session)

A commonly held view within the EA community is that it's incredibly important to start from thinking about what it really means to make a difference, before thinking about specific ways of doing so. It's hard to do the most good if we haven't tried to get a clearer picture of what doing good means, and as we saw in Week 2, clarifying our views here can be quite a complex task.

One of the core commitments of Effective Altruism is the ethical ideal of impartiality. Although in normal life we may reasonably have special obligations (e.g. to friends and family), in their altruistic efforts aspiring effective altruists strive to avoid privileging the interests of others based on factors such as space or time.
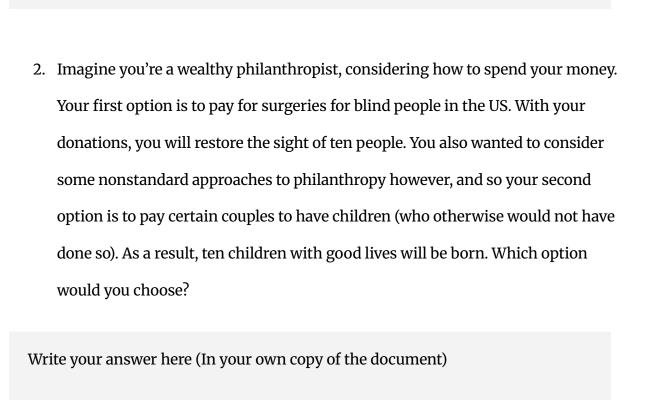
Longtermism posits that we should also avoid privileging the interests of individuals based on when they might live.

In this week's exercise, we'll be reflecting on some prompts to help you start considering what you think about this question, i.e. "Do the interests of people who are not alive yet matter as much as the interests of people living today?"

Spend a couple of minutes thinking through each prompt, and note down your thoughts – feel free to jot down uncertainties or open questions you have that seem relevant. We encourage you to note down your thought process, but feel free to simply report your intuitions and gut feelings.

1. Imagine you could save 100 people today by burying toxic waste that will, in 200 years, leak out and kill thousands (for the question, assume you know with an unrealistic level of certainty that thousands will die). Would you choose to save the 100 now and kill the thousands later? Does it make a difference whether the toxic waste leaks out 200 years from now or 2000?

Write your answer here (In your own copy of the document)

2. Imagine you're a wealthy philanthropist, considering how to spend your money. Your first option is to pay for surgeries for blind people in the US. With your donations, you will restore the sight of ten people. You also wanted to consider some nonstandard approaches to philanthropy however, and so your second option is to pay certain couples to have children (who otherwise would not have done so). As a result, ten children with good lives will be born. Which option would you choose?

Write your answer here (In your own copy of the document)

Did you consider what the implications of these interventions would be for people in the world other than the people specified in the question? If not, would that change your answer?

3. Imagine you donate enough money to the Against Malaria Foundation to save a life. Unfortunately, there's an administrative error with the currency transfer service you used, and AMF aren't able to use your money until 5 years after you donated. Public health experts expect malaria rates to remain high over the next 5 years, so AMF expects your donation will be just as impactful in 5 years time. Many of the lives that the Against Malaria Foundation saves are of children under 5, and so the life your money saves is of someone who hadn't been born yet when you donated.

    If you had known this at the time, would you have been any less excited about the donation?

Discussion Questions

1. What is longtermism?

2. What are the arguments for longtermism? Do they convince you?

3. What are some counterarguments or uncertainties you hold?

4. Are there tractable longtermist interventions other than reducing extinction/catastrophe risk?

Optional Materials

☐ [Greaves and MacAskill, The Case for Strong Longtermism](#) (1.5 hours)

☐ [80,000 Hours on Patient Longtermism](#) (5 mins)

☐ [Theron Pummer, How Important Is Population Ethics? (2014)](#) (20 mins)

☐ [Rational Animations on Longtermism](#) (6:11)

☐ [Will MacAskill – What We Owe the Future](#) (40:36)

Week 4: Existential Risk

"So if we drop the baton, succumbing to an existential catastrophe, we would fail our ancestors in a multitude of ways. We would fail to achieve the dreams they hoped for; we would betray the trust they placed in us, their heirs; and we would fail in any duty we had to pay forward the work they did for us. To neglect existential risk might thus be to wrong not only the people of the future, but the people of the past."

– Toby Ord

This week we'll cover the definition of an existential risk; examine why existential risks might be a moral priority; and explore why existential risks are so neglected by society.

Organization Spotlight

Future of Humanity Institute

The Future of Humanity Institute (FHI) is a multidisciplinary research institute working on big picture questions for human civilisation and exploring what can be done now to ensure a flourishing long-term future.

Currently, their four main research areas are:

- **Macrostrategy** – investigating which crucial considerations are shaping what is at stake for the future of humanity

- **Governance of AI** – understanding how geopolitics, governance structure, and strategic trends will affect the development of advanced artificial intelligence

- **AI Safety** – researching computer science techniques for building safer artificially intelligent systems

- **Biosecurity** – working with institutions around the world to reduce risks from especially dangerous pathogens

Core Materials

- *[The Precipice](#)*, Chapter 2 – Existential Risk (65 mins)
- *[The Precipice](#)*, Chapter 4 – Anthropogenic Risks (65 mins)

Discussion questions

Defining existential risk

1. Can someone give a definition of existential catastrophe and of existential risk?
2. Other than extinction, what other kinds of existential risk might there be?

Importance of reducing existential risk

3. So we've been talking about the idea that we and everyone we know could die in a catastrophe. That's pretty intense. How do you feel about it, what are your emotional reactions to it?

4. Can anyone explain Ord's reasoning about why an existential catastrophe is so much worse than other global catastrophes?

5. What do you think about Toby Ord's estimates of existential risk? What are the implications of existential risk being high or low?

6. Do you agree with Toby Ord that existential risks are neglected by society?

7. What are the best arguments that we shouldn't be this worried about existential risk?

Other

8. Do you think there are other existential risks that we haven't discussed?

9. Should we consider doing things that could do a huge amount of good, but don't have lots of supporting evidence, or should we do things that we have strong evidence do a lesser amount of good? How should we decide?

Optional Materials

- [Policy and research ideas to reduce existential risk – 80,000 Hours](#) (5 mins)
- ☐ [Reducing Global Catastrophic Biological Risks Problem Profile – 80,000 Hours](#) (60 mins)
- ☐ [Professor Stuart Russell on the flaws that make today's AI architecture unsafe & a new approach that could fix it](#)

More to explore

- [ ] *[The Precipice](#)* – Chapter 3 Natural Risks– *How big is the threat to humanity posed by asteroids and comets, supervolcanoes, stellar explosions, and other natural risks? (60 mins)*

- [ ] [An analysis and evaluation of methods currently used to quantify the likelihood of existential hazards](#) – *What are the different ways of evaluating existential risk, and which are the most robust? (70 mins)*

- [ ] [The Vulnerable World Hypothesis – Future of Humanity Institute](#) – *Scientific and technological progress might change people's capabilities or incentives in ways that would destabilize civilization. This paper introduces the concept of a vulnerable world: roughly, one in which there is some level of technological development at which civilization almost certainly gets devastated by default.* (45 mins)

- [ ] [Open until dangerous: the case for reforming research to reduce global catastrophic risk](#) (Video – 50 mins)

- [ ] [Dr Greg Lewis on COVID-19 & the importance of reducing global catastrophic biological risks](#) (150 mins for the podcast, 60 mins for transcript.)

Global governance and international peace

- [ ] [Ambassador Bonnie Jenkins on 8 years of combating WMD terrorism](#) – *an interview with Bonnie Jenkins, Ambassador at the U.S. Department of State under the Obama administration, where she worked for eight years as Coordinator for Threat Reduction Programs in the Bureau of International Security and Nonproliferation.* (Podcast – 1 hour 40 mins)

- ☐ [Why effective altruists should care about global governance](#) – *Because global catastrophic risks transcend national borders, we need new global solutions that our current systems of global governance struggle to deliver.* (Video – 20 mins)
- ☐ [Destined for War: Can America and China Escape Thucydides's Trap](#) (Book)

Climate change

- ☐ [Climate Change Problem Profile – 80,000 Hours](#) – *An analysis of the worst risks of climate change, and some of the most promising ways to reduce those risks.* (30 mins)
- ☐ [What can a technologist do about climate change?](#) – *A wide collection of technical projects to reduce the burning of fossil fuels.* (60 mins)

Nuclear security

- ☐ [Daniel Ellsberg on the creation of nuclear doomsday machines](#) – *Daniel Ellsberg on the institutional insanity that maintains large nuclear arsenals, and a practical plan for dismantling them* (Podcast – 2 hours 45 mins)
- ☐ [List of nuclear close calls – Wikipedia](#) – *A description of the thirteen events in human history so far that could have led to an unintended nuclear detonation* (5 mins)

Week 5: Emerging Technologies

One way to look for opportunities to accomplish as much good as possible is to ask "which developments might have an extremely large or irreversible impact on human civilisation?" During this week, we'll explore a few technological trends which might have relevance for existential risk. This week, understandably, can't cover all the major considerations for what the future will be like, but we aim to cover two key emerging technologies that might be less well known – transformative artificial intelligence and advances in biotechnology.

Organization Spotlight

[Centre for Security and Emerging Technology](#)

The Center for Security and Emerging Technology (CSET) is a policy research organization that produces data-driven research at the intersection of security and technology, providing nonpartisan analysis to the US policy community.

They are currently focusing on the effects of progress in artificial intelligence, advanced computing and biotechnology.

CSET is aiming to prepare the next generation of decision-makers to address the challenges and opportunities of emerging technologies. Their staff include renowned experts with experience directing intelligence and research operations at the National Security Council, the intelligence community and the Departments of Homeland Security, Defense and State.

Required Materials

Pick two of the below:

- ☐ *The Precipice* – Chapter 5 (pages 121–138) – Pandemics (25 min.)
- ☐ Risks from Great Power Conflicts – Video with transcript (28:18)
- ☐ Legal Priorities Project Research Agenda, 5.2.1–5.25 at p. 774, 5.2.1 Global Cooperation (15 mins)
- ☐ Reducing global catastrophic biological risks (GCBRs), stop at "How to Help" (35 mins)
- ☐ Learning to deal with dual use (5 mins)
- ☐ Gaps in international governance of dual–use research of concern (35 mins), (focus on pages 1–8)
- ☐ Common misconceptions about biological weapons (30 mins)

Exercise (30 mins)

Every day each of us makes judgments about the future in the face of uncertainty. Some of these judgments can have a huge impact on our lives, so it's really important that we make them as accurately as possible. But what can you do if you have limited information about the future? This week we'll practice making predictions, with the goal of honing your ability to make accurate judgments in uncertain situations.

The aim of the exercise is to help you become "well-calibrated." This means that when you say you're 50% confident, you're right about 50% of the time, not more, not less; when you say you're 90% confident, you're right about 90% of the time; and so on. The app you'll use contains thousands of questions - enough for many hours of calibration training - that will measure how accurate your predictions are and chart your improvement over time. Nobody is perfectly calibrated; in fact, most of us are overconfident. But [various studies](#) show that this kind of training can quickly improve the accuracy of your predictions.

Of course, most of the time we can't check the answers to the questions life presents us with, and the predictions we're trying to make in real life are aimed at complex events. The *Calibrate Your Judgement* tool helps you practice on simpler situations where the answer is already known, providing you with immediate feedback to help you improve.

[Exercise - Calibrate Your Judgement](#)

Discussion Questions

Past trends and transformative events

1. Imagine you're trying to do as much good as you can. Would you use a time machine to try to influence history? Do you think there's another time when you could have had more impact than now?

   - Let's assume that you're a similar kind of person to now (eg. an educated person living in a developed nation, but you don't have access to future technologies and knowledge)

   - Things you might want to bring up: agricultural revolution, US constitution, the industrial revolution, before/during/after WW1/2

   - Could an altruist in the past have had a greater impact if they knew the transformative events on the way? What would their best strategy have been to predict and influence these events?

2. Imagine you could speak to a 16th-century ruler of a country or nation who wanted to do as much good as they could. What would you advise them?

Future trends and transformative events

3. What important changes do you expect to happen in the future? Why?

4. What things do you think will happen in the next 200 years that could greatly impact the trajectory of human civilization?

5. Is influencing potentially transformative future events a promising approach for an altruist today? Why? What might change your mind?

Optional Materials

General Reading

Readings with an * have topics we can explore in more depth in focus weeks.

- ☐ [Global Catastrophic Risks Chapter 20 - Biotechnology and Biosecurity](#)*
  Biotechnological power is increasing exponentially, at a rate as fast or faster than that of Moore's law, as measured by the time needed to synthesise a certain sequence of DNA. This has important implications for biosecurity. (60 mins)

- ☐ [Efforts to Improve Accuracy in our Judgements and Forecasts - Open Philanthropy](#) (10 min.)

- ☐ [Some Background on Our Views Regarding Advanced Artificial Intelligence - Open Philanthropy Project](#) – *An explication of why there is a serious possibility that progress in artificial intelligence could precipitate a transition comparable to the Neolithic and Industrial revolutions.* (60 mins)*

- ☐ [Three wild speculations from amateur quantitative macrohistory](#) (10 min.)

- ☐ [What Failure Looks Like](#) (12 minutes) – *Two specific stories about what a very bad society-wide AI alignment failure could look like, which differ considerably from the classic "intelligence explosion" story*

Legal Reading

- ☐ [Jurisprudential Space Junk Treaties and New Technologies](#) – *skim to consider anecdotes on whether international law has any independent force*

- [ ] [Research agenda](), Legal Priorities Project, 5.1.8 at p. 67, Flexible and Clear Regulatory Approach

Global historical trends

- [ ] [How big a deal was the Industrial Revolution?]() (1hr. 20 mins)
- [ ] [Modeling the Human Trajectory – Open Philanthropy Project]() (30 mins)
- [ ] Books on macrohistory: [*Guns, Germs, and Steel*](), [*Global Economic History: A Very Short Introduction*](), or [*Sapiens*]()

Biosecurity

- [ ] [*Current Topics in Microbiology and Immunology – Volume 424, Chapter 7*]() – Does Biotechnology Pose New Catastrophic Risks? – *A description of the challenges of managing dual-use capabilities enabled by modern biotechnology.* (60 mins)
- [ ] ["Designer bugs": how the next pandemic might come from a lab]() – V0x – *what are superbugs, and why are they a possible existential risk? (25 mins)*
- [ ] [Explaining Our Bet on Sherlock Biosciences' Innovations in Viral Diagnostics – Open Philanthropy Project]() – *The Open Philanthropy Project report on their investment in Sherlock Biosciences to support the development of a diagnostic platform to quickly, easily, and inexpensively identify any human virus present in a sample.* (15 mins)
- [ ] [*Biosecurity Dilemmas*]() by Christian Enemark – *Introduction to the history and core dilemmas facing those working in biosecurity* (intro & conclusion: 20 mins).

Information hazards (similar to biosecurity)

[Information Hazards in biosecurity – Lewis et al., 2018](#) – *What are information hazards, how does it influence our idea of biosecurity, and what can we do about them?* (30 mins)

Artificial intelligence

- ☐ [What is artificial intelligence? Your AI questions, answered – Vox](#) (40 mins)
- ☐ [The new 30-person research team in DC investigating how emerging technologies could affect national security – 80,000 Hours](#) – *How might international security be altered if the impact of machine learning is similar in scope to that of electricity?* (Podcast – 2hr.)
- ☐ [AGI Safety from first principles](#) (1 hr 15 mins) – *one AI PhD student's take, from first principles, on the specific factors for the problem of aligning general AI*
- ☐ [Potential Risks from Advanced Artificial Intelligence: The Philanthropic Opportunity – Open Philanthropy Project](#) – *Why reducing risks from AI might be one of the most outstanding philanthropic opportunities.* (40 mins)
- ☐ [*Human Compatible: Artificial Intelligence and The Problem of Control*](#) (Book)
- ☐ [*The Alignment Problem: Machine Learning and Human Values*](#) (Book)

Other

- ☐ [Technology Roulette: Managing Loss of Control as Many Militaries Pursue Technological Superiority – Centre for a New American Security](#) – *An argument for how advances in military technology (including but not limited to AI) can impede relevant decision making and create risk, thus demanding greater attention by the national security establishment.* (60 mins)

- [ ] [Big nanotech: towards post-industrial manufacturing - The Guardian](#) – *an explanation of how atomically precise manufacturing could displace industrial production technologies and bring radical improvements in production cost, scope, and resource efficiency.* (10 mins)

- [ ] [AlphaGo - The Movie - DeepMind](#) – *A documentary exploring what artificial intelligence can reveal about the 3000-year-old game of Go, and what that can teach us about the future potential of artificial intelligence.* (Video – 1hr. 30 mins)

- [ ] [The Artificial Intelligence Revolution: Part 1](#) – *A fun and interesting exploration of artificial intelligence by the popular blogger Tim Urban.* (45 mins)

- [ ] [The Future of Surveillance](#) – *An exploration of ways in which the future of surveillance could be bad, and an investigation into accountable, privacy preserving surveillance protocols.* (Video – 15 mins)

Week 6: Focus Week 1

Here, we'll discuss whatever topic we chose the previous week. Previously curated topics are

beneath Week 8, but let us know if there are ones not on the list you'd like to discuss!

Week 7: Focus Week 2

Here, we'll discuss whatever topic we chose the previous week. Previously curated topics are

beneath Week 8, but let us know if there are ones not on the list you'd like to discuss!

Week 8: Putting it into Practice

Core Materials

- ☐ [Three Key Career Stages](#) (10 mins)

- ☐ [You have more than one goal, and that's fine](#) (5 mins)

Discussion Questions

1. What were your plans before the reading group?

2. Has the reading group made you consider changing your plans or help you rule out any career options you were considering?

3. More generally, how useful do you think legal education/skills are for doing the most good in any plausible cause area?

4. What advice would you have for someone considering law school as a path to impact?

Optional Materials

- ☐ [Law School: Why and When?](#) (57 mins) *Some of the considerations are U.S. specific, and aren't as relevant to Canadian law students, but some others are spot on. Worth the read and send to friends considering law school.*

- ☐ [80,000 Hours Career Planning Guide – Parts 3–8](#) (1.5 hrs.)

- ☐ [Evidence-based advice on how to be successful in any job – 80,000 Hours](#) (45 mins)

- ☐ [How useful is long-term career planning? – 80,000 Hours](#) (10 mins)

- ☐ [Advice on how to read our advice – 80,000 Hours](#) (10 mins)

Action Item

- ☐ [Visit 80,000 Hours' Job Board](#)

- ☐ [Apply for 80,000 Hours' 1-on-1 Advising](#)

Focus Week Topics

Improving Institutional Decision-Making

If you're curious about working in Legal Rights, please reach out to your group facilitator for further information.

Core Materials

- ☐ [Improving institutional decision-making | Jess Whittlestone | EA Global: London 2017](#) (video – 28:46)
- ☐ [80,000 Hours on Improving Institutional Decision-making](#) (article – 30 mins)

Exercise: Probabilistic Fallacies (40 mins)

Have you ever felt absolutely certain that you were right about something, only to discover that you were wrong? If so, there's a good chance that you misinterpreted the evidence. This mini-course will train you to think more clearly about evidence and how it works, so you can come to the right conclusions when it counts.

- ☐ Work through the mini-course and get better at interpreting evidence [here](#).

Discussion Questions

1. Do you think there is a way to disentangle institutional design from political power? (i.e. – will those with the power ever choose incentives likely to conflict with their retention of power?)

2. Is the cost-benefit analysis (or similar) framework applied to broad legal mandates sensitive to traditional EA concerns? Which ones? What are the easiest ways it could be made more sensitive? Through the courts? Through new regulations?

3. What are some externalities (positive or negative) that societies fail to internalize? Are there ways of internalizing them without wielding political power?

4. What would a consequence- or utility-based legal rule look like? Can you think of a reliable means for enforcing such a rule, absent political will?

5. What would an effective longtermist international institution look like? What would be needed for it to come about? How would we gauge its effectiveness and what incentives would dictate its behaviour?

6. What role does legal interpretation have to play in bringing about positive social/legal change? How does it compare to legislative or cultural change? Are there instances where legal interpretation is more or less tractable than those?

Optional Reading

☐ [Legal Priorities Project Research Agenda](), "Institutional Design" (pp. 83–98); Flexible Constitutions (6.1.3 at p. 84) (30 mins)

☐ [Longtermist Institutional Reform]() (30 mins)

☐ [The Question of Evidence]() (35 mins) – *Test your ability to form accurate beliefs.*

- ☐ [Efforts to Improve the Accuracy of Our Judgments and Forecasts - Open Philanthropy Project](#) (15 mins)

- ☐ [How To Replace Our Broken Voting System](#) (15 min – or you can [watch the talk here](#))

- ☐ [European Commission (2020): Technology and Democracy: Understanding the influence of online technologies on political behaviour and decision-making](#)

Career Resources

- ☐ [80,000 Hours' Improving institutional decision making](#) (45 mins)

New Holders of Legal Rights

If you're curious about working in Legal Rights, please reach out to your group facilitator for further information.

Core Materials

- ☐ [Kurki, V. A. J. (2019). "A Short History of the Rights-holding Person." Oxford University Press.](#) (40 mins)

- ☐ [Personhood Initiatives](#) (27:48)

Discussion Questions

1. Effective Altruists tend to think in terms of interests and trade-offs rather than a fixed set of legal rights. The law does the opposite -- how might a rights framework better promote welfare than it does currently?

2. What are the principal philosophical theories justifying the classes of people who enjoy legal rights today?

3. What practical considerations go into granting legal rights? To what extent do rights-holders need to be able to communicate their interests to have their rights represented?

4. When rights-holders can't communicate their interests at all, who gets to represent them? On what basis? Under what circumstances?

5. Who poses the biggest threat to unrepresented interests? When does it make sense to sue them? Under what theories?

Optional Materials

- ☐ [Institutions for Future generations](#) – Video (31:47)

- ☐ [Legal personhood and the positive rights of animals](#) (30 mins) (esp. Sections 4.4, 5)

- ☐ [Cass Sunstein, "Standing for Animals,"](#) Chicago Unbound, Public Law and Legal Theory Working Paper No. 06 (1999).

- ☐ [Why Lawsuits over Misleading Food Labels are Surging (NYT)](#) (10 mins)

- ☐ [Longtermism and Animal Advocacy](#) (15 mins)

- ☐ [Research – All-Party Parliamentary Group for Future Generations](#) (5 min – 2hr.)

- ☐ [Defending the Future: Intergenerational Equity in Climate Litigation](#), 32 Geo. Envtl. L. Rev. 569 (2020).

Regulating Artificial Intelligence

If you're curious about working in Artificial Intelligence, please reach out to your group facilitator for further information.

Core Materials

- ☐ [Potential Risks from Advanced Artificial Intelligence: The Philanthropic Opportunity – Open Philanthropy Project](#) – *Why reducing risks from AI might be one of the most outstanding philanthropic opportunities.* (40 mins)

- ☐ [*The Precipice* – Chapter 5 (pages 138-152) – Unaligned Artificial Intelligence](#) (25 min.)

Discussion Questions

1. How should AI and machine learning be defined for legal purposes? What are the consequences of different definitions? Think about different dimensions, e.g.,

    a. technology-based regulation vs. risk-based regulation

    b. regulation vs. co-regulation vs. self-regulation

    c. hard law vs. soft law (e.g., international standards)

    d. international regulation vs. national regulation

2. Does the regulation of the internet provide a helpful analogy to the regulation of AI?

a. The early internet was largely unregulated in the United States (e.g. no enforced sales tax on ecommerce; no speech liability for platform providers; no treatment of ISPs as utility providers) - what are the most compelling rationales for treating AI differently?

3. Much of the concern around AI considers a "hard takeoff" scenario for AGI, what if any current legal tools would help minimize a hard take off happening and/or going badly?

4. How open to competition should the AI development process be? Are their risks in trying to define AI research and development broadly and regulate it? What does regulation of non-public advances even look like? How is this different than arms control?

5. Could the recognition of rights for AI create a moral hazard for designers or parties who would otherwise be held accountable?

6. Would a windfall clause be legal in Canada? Can you think of other ways to distribute the benefits from AI?

Optional Materials

Generally

☐ [Research Agenda](), Legal Priorities Project, Sections 4.2–4.4, pp. 46–55

☐ [Public policy and superintelligent AI: A vector field approach]() (1 hour)

☐ [Should AI governance be centralised? – Design lessons from history]() (1 hour)

- [ ] [Why Responsible AI Development Needs Cooperation on Safety](#) (7 mins)

- [ ] [Research Agenda](#), Legal Priorities Project, Section 4–4.1, pp. 35–46 (20 mins)

Misuse risks (security)

- [ ] [Malicious Use of AI: Forecasting, Prevention, and Mitigation](#) (2 hours)

Accident risks (safety)

- [ ] [Regulating AI Systems: Risks, Challenges, Competencies, and Strategies](#) (1.5 hours)

- [ ] [Concrete problems in AI safety](#) (1.5 hours)

- [ ] [Regulatory markets for AI safety](#) (1.5 hours)

- [ ] [Toward trustworthy AI development: Mechanisms for supporting verifiable claims](#) (1.5 hours)

Structural risks/opportunities

- [ ] [The Windfall Clause: Distributing the Benefits of AI for the Common Good](#) (1 hour)

Career resources

- [ ] [Guide to working in AI policy and strategy](#), 80,000 Hours

- [ ] [The case for building expertise for US AI policy and how](#), 80,000 Hours

Animal Law

If you're curious about working in Artificial Intelligence, please reach out to your group facilitator for further information.

Core Materials

- ☐ [The Case Against Speciesism - Centre for Reducing Suffering](#) (10 mins)

- ☐ [The Relevance of Wild Animal Suffering](#) (10 mins)

- ☐ Choose one optional reading for a deep dive

Discussion Questions

1. Are personhood and "animal rights" useful concepts for legally advancing animals' interests?

2. Are current legal protections for animals inherently tied to human concerns or popular initiatives?

3. What have been the main legal obstacles to progress on animal welfare? What accounts for animal advocates' recent legal successes against the meat industry? Is there a limit to what can be defended under the current law?

Optional Materials

- ☐ [The fight over cage-free eggs and bacon in California](#) (15 mins)

- ☐ [Plant-Based Labelling Battle Heats Up](#) (10 mins)

- ☐ [Quantifying Americans' Valuation of Animal Suffering talk by Scott Weathers](#) (30 mins)

- ☐ [Why and How to Value Nonhuman Animals in Cost-Benefit Analyses](#) by Andrew Stawasz, (3-9; 34-70)

- ☐ Blattner, C. (2019). The recognition of animal sentience by the law. Journal of Animal Ethics. (via Slack)

- ☐ [Standing for Animals (With Notes on Animal Rights)](#), 47 U.C.L.A. Law Review 1333 (2000) (edited version).

- ☐ *McKiver v. Murphy-Brown, LLC*, No. 19-1019, at *67 (4th Cir. Nov. 19, 2020) (Wilkinson's Concurrence only) (via Slack)

- ☐ [The Animal Welfare Act: Background and Selected Animal Welfare Legislation](#)

- ☐ *National Meat Association v. Harris*, 132 S. Ct. 965 (2012) (addressing pre-emption of California's non-ambulatory livestock rule) (via Slack)

- ☐ *Jones v. Butz*, 374 F.Supp 1284 (1974) (challenging religious basis for kosher/halal slaughter's inclusion in the humane slaughter act) (via Slack)

Career Resources

- ☐ [80,000 Hours' Resources on Factory Farming](#) (70 mins)
- ☐ [Jeffray Behr's Career Guide for Ending Factory Farming](#) (29 mins)
- ☐ [80,000 Hours' Job Board on Factory Farming](#)