Existential Advocacy

John Bliss

LPP WORKING PAPER SERIES N° 4-2022
Existential Advocacy

John Bliss

September 10, 2022

Preliminary Draft

ABSTRACT

Lawyers have played a central role in a wide range of social movements. How is this tradition of legal activism being applied in the emerging movement to protect the future of humanity? This movement is motivated by growing evidence that humanity is entering a new age of “existential risk,” which refers to events that would entirely foreclose a meaningful existence for future generations, e.g. via human extinction of permanent dystopia. Technological developments in synthetic biology, artificial intelligence, and other fields have brought these risks out of science fiction and into serious academic inquiry and political concern. Leading scholars in this field estimate a roughly one-in-six chance of existential catastrophe over the next century. Awareness of risks on this scale has recently sparked the founding of dozens of non-profit organizations and the early stirrings of a transnational social movement. This article presents the first empirical study of the nascent community of legal activists within this movement. How do these lawyers and policy advocates address such an
abstract and probabilistic topic as existential risk? And how do they represent the multitudes of future generations, a population that could be extraordinarily large yet entirely voiceless? And what can they learn from the interdisciplinary literature on law and social change?

Drawing on a year-long qualitative study including semi-structured interviews with the leading legal advocates and ethnographic observations embedded in the central legal organization of this movement, the article describes in empirical detail the distinct model of social-change lawyering practiced in this movement. This model, which the article labels “prioritist advocacy,” is marked by a rigorous, evidence-based effort to maximize impact on high-priority goals (e.g., existential risk). To support the realization of this model, legal activists in this field have developed a culture of scientific truth-seeking norms, including efforts to reduce cognitive biases that inhibit recognition of existential threats and moral consideration of future generations. The article concludes by considering how the prioritist model may need to adapt as this movement scales up and pursues more direct and higher profile legal interventions.
# CONTENTS

INTRODUCTION .................................................................................................................. 4

I. BACKGROUND: ............................................................................................................ 12
   A. THE PROBLEM OF EXISTENTIAL RISK ................................................................. 12
   B. SOCIO-LEGAL LITERATURE ON LAW AND SOCIAL CHANGE .............................. 18

II. RESEARCH DESIGN ....................................................................................................... 21

III. THE PRIORITIST THEORY OF EFFICACY ................................................................. 24
   A. SELECTING GOALS ................................................................................................. 24
   B. SELECTING STRATEGIES ....................................................................................... 28

IV. THE PRIORITIST THEORY OF ACCOUNTABILITY ...................................................... 34

V. THE CULTURE OF PRIORITIST ADVOCACY ............................................................... 40
   A. NORMALIZING UNCERTAINTY .............................................................................. 40
   B. EMBRACING DISSENT .......................................................................................... 41
   C. LIMITING SYSTEM-1 DELIBERATION ................................................................ 43
   D. LIMITING IDENTIFICATION ............................................................................... 45

VI. DISCUSSION .................................................................................................................. 48
INTRODUCTION

“In some ways [the movement to mitigate existential risk] is the most inclusive movement, because it’s trying to preserve everything. It sort of sits above all social movements...as a continuation of social justice. I think it’s really vitally important to every person in the world today and every single future person....”

“[When presenting existential risk to legal and political actors, they often respond] ‘How can I think about this when we’ve got so many crocodiles closer to the boat?’”

“[When advocating for the mitigation of existential risk] if the core aggrieved group is future generations, they don’t get to chime in because: no time travel.”

— Excerpts from legal activists interviewed in this study

Imagine a pandemic 1,000 times more deadly than Covid-19. Or imagine artificial intelligence so powerful that it could end or enslave the human species. Threats of this scale—known as “existential risks” for their ability to permanently bring an end to any meaningful existence of humanity—are the subject of a growing movement of legal activism. This article provides the first empirical examination of this nascent but legal community. Drawing on a multi-method study, I examine how these advocates conceive of social-change lawyering in the unique context of existential risk. What does it mean to be advocate around an issue, which, as described in the interviews excerpted above, operates on an unimaginably large scale (affecting “every person in the world and every single future person”) but is difficult to address due to a host of cognitive and political factors directing our focus to issues that are more certain and immediate (the “crocodiles closer to the boat”)? Moreover, how can advocates faithfully represent future generations, a vast and entirely voiceless constituency who are silent stakeholders in decisions we make today (and who cannot “chime in” due to the lack of “time travel”)? This is lawyering in an entirely new age. For the first time in history, humanity appears to be developing tools

---

1 See generally, Nick Bostrom, Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards, 9 J. EVOLUTION AND TECH. (2002); TOBY ORD, THE PRECIPICE: EXISTENTIAL RISK AND THE FUTURE OF HUMANITY 5 (2020) (defining existential threats as “risks that threaten the destruction of humanity’s longterm potential,” which is “most obvious” with human extinction but also includes other irreversible future trajectories of immense suffering).
capable of not only mass but total destruction. As the list of such technologies continues to grow, how do legal activists conceive of their role within a larger portfolio of social-change strategies?

The notion of catastrophes that would irreversibly terminate any meaningful human future has historically been the stuff of science fiction and “disaster films.” But, over the past two decades, the study of catastrophes on this scale has grown into a serious field of empirical and theoretical inquiry. In his comprehensive treatment of this topic, The Precipice: Existential Risk and the Future of Humanity (2020), the philosopher Toby Ord estimated that the next century brings a one-in-six chance of existential catastrophe. This estimate takes into account scientific and theoretical investigations of a variety of threats. Ord collected these perspectives in his position as a senior fellow at Oxford University’s Future of Humanity Institute, where the field of existential risk studies originated. For the younger generation today, as indicated in surveys of people under 25 years old, this estimate of existential risk might not be especially surprising, given their sense that “humanity

---


3 The term existential risk was coined in Nick Bostrom in 2002, supra note 1. See generally, NICK BOSTROM AND MILAN M. CIRKOVIC, GLOBAL CATASTROPHIC RISKS; ANNETTE BAIER, RIGHTS OF PAST AND FUTURE PERSONS; Nick Bostrom, Existential Risk Prevention as a Global Priority; JOHN LESLIE, THE END OF THE WORLD: THE SCIENCE AND ETHICS OF HUMAN EXTINCTION.

is doomed.” But this popular sense of doom is usually associated with climate change and other events that would devastate large populations over a number of generations, and thus are worthy of great concern, but are projected to be very unlikely to foreclose the long-term future of humanity. The threats that are far more likely to have impact on the existential scale are thought to arise not from the more familiar issues of climate change, asteroids, or nuclear weapons, but rather from new and emerging technologies, and the sense that we have only just started what will become a growing list of abilities to produce existential catastrophes. As discussed in greater detail in Part I (infra), this includes developments in synthetic biology, with new abilities to modify pathogens, raising the risk of engineered super pandemics. Another leading concern relates to developments in artificial intelligence, with expert opinions increasingly suggesting that these systems will eventually (though there is debate about how soon) be able to surpass human reasoning and control. Moreover, transformative artificial intelligence may radically accelerate scientific and technological development, creating a strange new world with existential threats that cannot currently be imagined.

The problem of existential risk is made more difficult by the global coordination that would be required to contain threats of this magnitude: An existential catastrophe arising in one jurisdiction would end humanity in all jurisdictions. Thus, developing comprehensive regulatory responses (e.g. preventing research that involves engineering extremely dangerous pathogens) in some but not all jurisdictions may fall entirely short. Moreover, by definition, we have no experience with events that terminate humanity. We cannot afford a single failure to prevent


6 See, e.g., ORD, supra note 1 (estimating the existential risk associated with both nuclear war and climate change at 1/1,000 over the next century, but offering far higher estimates for unaligned artificial intelligence (1 in 10), engineered pandemics (1 in 30), and unforeseen anthropogenic risks (1 in 30)).

7 ORD, supra note 1 at (observing that the absence of strong multilateral mechanisms of global governance presents a formidable challenge for regulating these risks in the nearly 200 countries of the world).
such a catastrophe, nor we can we rely on a reactive trial-and-error approach to designing our responses. These observations have led some leading minds in this field to suggest that we are entering a new historical era, walking along a “precipice” in humanity’s “most important century” as we make our initial encounter with existential threats. As the issue is often framed in this field, humanity could either be approaching its end or, if we are able to survive the advent of technologies capable of bringing about existential catastrophes, and if we are able to establish global regulatory systems to assure that any new existential threats that arise would be contained as well, we could be still in the early infancy of humanity with an extraordinarily long arc of (hopefully good) experience ahead of us.

What started as a conversation among philosophers and scientists has now spread to a larger movement that not only engages in research but aims to actually intervene in existential threats. Non-profit organizations have proliferated in this field, as well a new wave of student-led “existential risks initiatives” at leading universities, including Stanford, Cambridge, Harvard, and MIT. The Stanford Existential Risks Initiative recently reported over 1,000 attendees from over 50 countries at the second meeting of their annual conference. The field has seen extraordinary fundraising in recent years, with the current pledged donations in the

---

8 Nick Bostrom, supra note 1 (“Our approach to existential risks cannot be one of trial-and-error. There is no opportunity to learn from errors. The reactive approach—see what happens, limit damages, and learn from experience—is unworkable. Rather, we must take a proactive approach. This requires foresight to anticipate new types of threats and a willingness to take decisive preventive action and to bear the costs (moral and economic) of such actions.”).


10 Organizations established around Oxford University include the Future of Humanity Institute, the Global Priorities Institute, and the Centre for Effective Altruism. Nearby Cambridge University is home to the Centre for the Study of Existential Risk. In the U.S., organizations working on existential risk include the Future of Life Institute and the Global Catastrophic Risk Institute. Grantmaking organizations include Open Philanthropy, Longview Philanthropy, Future Fund, the Berkeley Existential Risk Initiative.

11 These student-led initiatives involve a range of activities including summer fellowships, conferences, and reading groups.

tens of billions of dollars, which surpasses many of the leading philanthropic foundations of the world. In the past few years, this field has turned to the legal and political advocacy that is the subject of this article. This includes new efforts to encourage members of the existential risk community to run for U.S. Congress and other public office. The primary donor behind these efforts has recently stated an intention to contribute up to one billion dollars in the 2024 U.S. election cycle in service of mitigating existential risk (i.e., to support candidates from the existential risk community and oppose candidates who are believed to greatly exacerbate existential threats).

In just the past three years, this field has started to explore the role of law and lawyers within this movement. There are some signs that the legal profession might be receptive to working on existential risk. The two most cited scholars in U.S. legal academia have written books on large-scale catastrophes, although neither connect their work to the Oxford-based literature on existential threats. A recent international survey found that legal scholars generally believe that legal action taken today can impact existential risk and the long-term future. This optimism may be well placed, as the participants in this study highlight a new landscape of promising legal interventions in this field, spanning from local efforts to prevent

---

13 See Benjamin Todd, Is Effective Altruism Growing? An Update on the Stock of Funding vs. People, 80,000 Hours (July 28, 2021), https://80000hours.org/2021/07/effective-altruism-growing (estimating that 46 billion dollars have been pledged to effective altruist causes, and that this figure has been growing at 37% per year); Naina Bajekal, Want to Do More Good? This Movement Might Have the Answer, Time (Aug. 10, 2022, 7:00 AM), https://time.com/6204627/effective-altruism-longtermism-william-macaskill-interview (noting that the funding committed to Effective Altruism far exceeds that raised by the Ford Foundation (roughly $16 billion) and the Rockefeller Foundation (roughly $6 billion).


specific threats (e.g., seeking an injunction to prevent a laboratory from conducting gain-of-function research) to far-reaching efforts to establish the legal interests of future generations (e.g., judicial recognition of personhood, standing, or a right to life). As researchers in this field have reported, future generations are now referenced in 81 national constitutions, in addition to a host of domestic law and international agreements, and some courts have recently shown an unprecedented willingness to enforce these provisions in the context of climate change litigation. Perhaps the most striking example is the 2021 German Constitutional Court decision striking down a national climate law, while citing “intertemporal guarantees of freedom” and a “special duty of care...for the benefit of future generations.” In some jurisdictions, new legislation and parliamentary groups have arisen with a focus on protecting future generations. One of the most remarkable ongoing developments is taking place at the United Nations where the Secretary General’s Office issued a report on threats to future generations, including references to the existential risk literature, and is planning a 2023 “Summit of the Future.” The Summit is expected to bring together heads of state from around the world to introduce a UN Declaration on Future Generations, a United Nations

---


20 See Neubauer v. Germany, Bundesverfassungsgericht [BVerfGE] [Federal Constitutional Court of Germany] [GCC], 1 BvR 2656/18, 78/20, 96/20, and 288/20, para. 266, Mar. 24, 2020 (....); see also State of the Netherlands v. Urgenda, HR 20 december 2019, HAZA C/09/00456689, Urgenda/State of the Netherlands) (requiring emissions reductions on the grounds of right to life and right to respect for private and family life from Articles 2 and 8 of the European Convention on Human Rights and well as the UNFCC provision to “protect the climate system for the benefit of present and future generations of humankind”); Hague District Court 2021 (applying a “standard of care” as a human rights obligation that applies to future generations). Cf. Juliana v. United States, 947 F.3d 1159, 1175 (9th Cir. 2020) (finding that a group of twenty-one young plaintiffs lacked Article III standing for lack of redressability and dismissing all claims).

21 *Id.*

Special Envoy for Future Generations, and a United Nations Futures Lab among other mechanisms to protect the long-term future (if approved by member states).²³

In order to gain an in-depth understanding of the legal activism that is developing in response to existential risk, this project draws on a multi-method qualitative research design consisting of ethnography, semi-structured interviews (n=53), and a systematic review of online materials. The primary site for the study was the sole legal organization of the existential risk community, the Legal Priorities Project (hereinafter “LPP”), which permitted me to interview their members and observe their daily operations over a five month period. Founded in 2020 as a 501(c)(3) by students and affiliates at Harvard Law School, LPP is a globally diverse organization—the membership has a clear tilt toward Europe and the U.S., but, over the period of ethnography, the roughly twelve members who attended LPP’s internal all-hands meetings included nationalities (and legal training) from Africa, Australia, Central America, Continental Europe, North America, and South America.²⁴ They operate remotely with occasional in-person meetings. Thus, the study was largely conducted via video-conferencing calls in addition to in-person visits to key sites (e.g., Oxford, London, Geneva, and Cambridge, Massachusetts).

Since its inception, LPP’s primary focus has been research. They pursue empirical projects that aim to guide later strategic decisions about the use of law. They have also emphasized educational outreach in the legal profession, including summer fellowships, institutes, and a speaker series. Over the past year, they have increasingly engaged in policy advising on matters relating to their legal expertise. They also recently formed legal strategies team, hiring a Costa Rican human rights attorney, a Dutch judge, and an American public interest lawyer.

The primary finding explored in this article is that these legal activists have developed a distinct model of social-change lawyering, which I label “prioritist advocacy.” This model departs from other legal campaigns examined in the empirical literature on law and social change. The participants in this study have developed a systematic process for prioritizing their goals and strategies. Even the question of what issues they should seek to address is evaluated through their prioritization formula. This approach is derived from theories of “effective altruism” that have


²⁴ It should be noted that this degree of geographic diversity is perhaps uncommon among organizations in this field, although LPP is not alone in seeking to diversify the geography, race, gender and other identities represented in the movement. This issue is discussed infra Part IV.
never before been applied to legal practice.\textsuperscript{25} In short, prioritist advocacy begins with first-principle moral commitments, e.g. enhancing overall human well-being, reducing extreme suffering, or promoting justice. It then applies a specific set of criteria to select causes where advocates believe they can have the greatest impact on their moral commitments. At the moment, mitigating existential risk appears to be the most favored collection of cause areas among the prioritist advocates, although a willingness to change this priority upon new evidence is a key commitment under this methodology. The prioritists then run nearly all strategic decisions through an analysis to determine which option is expected to have the greatest impact on their chosen goals (e.g. reducing existential risk). As observed in this study, this model is supported by a daily culture of reinforcing scientific, truth-seeking norms, including (1) embracing epistemic humility and normalizing uncertainty, (2) fostering a warm, inclusive, and empathetic atmosphere of “supportive dissent,” and (3) limiting cognitive biases that would interfere with their focus on existential risk and maximizing impact. These scientific commitments are paired with a strong sense of accountability to future generations—although this raises difficult points of tension about how to represent a future population that is voiceless and how to incorporate a broader set of current-person voices in strategic decision-making processes. The full model is depicted in Figure 1.

\textbf{Figure 1.} The Prioritist Advocacy Model

\textsuperscript{25} See discussion of effective altruism infra Part III.
Before detailing these empirical findings, Part I provides background regarding the issue of existential risk and the socio-legal literature on lawyers within movements for social change. Part II reviews the methods of this study. Parts III through V present the empirical description of prioritist advocacy, beginning with theories of efficacy and accountability, and then discussing the movement’s underlying cultural commitments. Part VI concludes the article with an assessment of prioritist advocacy and recommendations for adapting this model as the movement around existential risk scales up and pursues more direct and higher profile legal interventions.

I. BACKGROUND:

A. The Problem of Existential Risk

Humanity only recently discovered the ability to destroy itself. Professor Nick Bostrom, the Oxford philosopher who coined the term “existential risk,” notes that until the advent of nuclear weapons there were “probably no significant existential risks in human history…and certainly none that it was within our power to do something about.” Even nuclear weapons are considered unlikely to foreclose a meaningful future, although the newer and more powerful generation of these weapons, if deployed around the globe on an enormous scale, could theoretically end humanity through an extreme “nuclear winter” scenario. Researchers in the emerging field of existential risk studies have now spent over two decades analyzing a wide range of conceivable threats on the scale of human extinction or otherwise irreversible destruction of human potential. This has led many in this field to the

---

26 See THOMAS MOYNIHAN, X-RISK: HOW HUMANITY DISCOVERED ITS OWN EXTINCTION (2020) (discussing the long-running history of religious traditions of apocalyptic prophecy, which portend a “sense of an ending,” but explaining that existential risk, and human extinction in particular, is a “comparatively novel idea” involving an “ending of sense,” a more complete ending of human experience).


29 See Bostrom, supra note (defining existential risks as threats to “either annihilate Earth-originating intelligent life or permanently and drastically curtail its potential”); ORD, supra note 1 (defining existential risk to include extinction, locked-in totalitarianism (“world in chains”), and irreversible collapse of civilization “where humanity across the globe loses civilization, a world without writing, cities, law, or any trappings of civilization.”). See generally, Policy Idea Database, CNTR. STUDY EXISTENTIAL RISK, https://www.gcrpolicy.com/ideas (last visited Aug. 15, 2022) (showing a total of 376 publications in the field of existential and catastrophic risk studies with an increase after 2017 to 75 to 82 publications per year); see also Rumtin Sepasspour, Release of
conclusion that existential risk should now be assigned a significant probability. Ord, who offered the 1/6 estimate for the next century, refers to our new era as “the Precipice,” defined by our initial meeting with threats of this magnitude. Ord’s optimistic framing of this era is that it will (hopefully) be remembered as the time when humanity “opened its eyes” to existential risk and “guaranteed a long and flourishing future” through transformations in our legal, political, normative, and cognitive frameworks.

While climate change and nuclear weapons are the most familiar catastrophic threats in our public discourse, Ord assigns these categories only a 1/1000 chance of reaching the scale of irreversibly destroying the human future over the next century. Existential threats from natural sources are similarly considered quite unlikely to materialize in the near future. Although asteroids loom large in fictional accounts, as well in prehistory as the source of the Cretaceous–Paleogene Extinction, this category is now well studied. By looking to the skies and to the earth (via the geological record), scientists have shown that such events occur infrequently and thus appear to pose a relatively slight danger in the near term. Ord suggests a dramatic contrast between the probability of natural and anthropogenic existential risks. For example, he assigns a 1/10,000 probability that natural pandemics will bring about an existential catastrophe over the next century, while engineered


30 ORD, supra note 1.
31 Id., at 31.
32 ORD, supra note 1 (noting that nuclear weapons and climate change “awoke us to the possibilities of destroying ourselves,” and noting that the threat of climate change could grow to existential proportions with a runaway greenhouse effect that boils the oceans or sets off a cascade of ecosystem failures, although experts tend to suggest that such total destruction scenarios are unlikely). Existential risk is often viewed as a subset of the broader category of global catastrophic risk, which threaten widespread and even global harm but may lack the “terminal” intensity of risks on the existential scale. See BOSTROM AND CIRKOVIC, supra note (defining global catastrophic risk as “a catastrophe that [would cause] 10 million fatalities or 10 trillion dollars worth of economic loss”).
33 Adam Mann, Odds of Death by Asteroid? Lower Than Plan Crash, Higher Than Lightning, Wired (Feb. 15, 2013, 2:39 PM), https://www.wired.com/2013/02/asteroid-odds (reporting on the risk table by NASA’s Near Earth Object program that reports on the “likelihood of impact for the next 100 years” of known asteroids).
34 See BOSTROM AND CIRKOVIC, supra note.
pandemics are assigned a probability of 1/30. Moreover, engineered pandemics are part of a growing category of concerns relating to synthetic biology, including bioweapons, pathogens escaping laboratories by accident or on purpose, or “information hazards” where details about how to create dangerous biological materials are not kept confidential. More speculatively, other frontiers of scientific endeavor could also pose existential threats, such as bringing back unpredictable materials from space exploration or running “radical scientific experiments” with unknown risks that could theoretically reach the scale of destroying life on Earth and beyond.

But scholars in this field tend to agree that the largest single category of existential risk, by a wide margin, relates to the development of transformative artificial intelligence (“AI”), estimated by Ord as a 1/10 existential threat over the next century. With new developments in deep learning, a growing chorus of AI experts now believe that it will be possible to develop AI systems that far surpass human intelligence in many or all respects. There is less consensus about when such technologies would be developed, although many experts believe that it is only a matter of decades or sooner. Such systems could raise existential risks if deployed maliciously by bad actors, or unsafely by actors who are in a race to develop the most powerful AI systems and deploy them quickly and expansively, applying them to increasingly important areas of life. This race dynamic appears to be emerging not only in the private sector but also among national governments who see enormous economic and military value in this technology, as described in the Chinese political leadership’s 2030 goal to achieve “AI dominance.”

---

35 See Ord, supra note 1 (reviewing the long history of the use of disease as a weapon, as well as examples of lab escapes and information hazards including the publication of the smallpox genome).

36 Id. (discussing “radical science experiments,” and citing the example of the advent of nuclear weapons when some scientists theorized that the first detonation would ignite Earth’s atmosphere and set off an existential catastrophe).

37 See, Data of Artificial Intelligence, https://www.metaculus.com/questions/5121/date-of-artificial-general-intelligence/ (reporting on a monthly basis the average expert estimate of the arrival of artificial general intelligence, which has wavered in recent months between the years 2034 and 2059, but noting that roughly 50% of AI researchers believe that such technology will not arrive until after the year 2100).


is that it may be difficult or impossible, even with our best efforts, to align these systems with human values and interests—to specify reward functions for superintelligent systems such that they reliably respect, at a bare minimum, human life and our interest in non-dystopic futures. Current concerns over algorithmic bias may be a warning shot of the challenges we will face when seeking to align transformative AI systems with good human values—in addition to the challenge of determining what good human values are and who should get to decide this. Ord and others also note that transformative AI may have both positive and negative effects on other existential threats, e.g. helping us develop tools to help with problems like climate change but rapidly accelerating the scientific advances that give rise to new threats on an existential scale.

Finally, the most uncertain category is unforeseen risks, particularly within the framework of Bostrom’s “vulnerable world hypothesis.” Bostrom posits that humanity’s never-ending practice of drawing out new inventions from the metaphorical “urn of creativity” may eventually yield a “black ball” technology, that is, a technology that “invariably or by default destroys the civilization that invents it.” An example could be a device with the power of nuclear weapons but an ease of assembly that requires only a commercially available 3D printer or the equipment in a typical garage. It is perhaps only by good fortune that humanity has not yet discovered a black ball, and it may be only a matter of time before we do. If we achieve transformative AI that greatly accelerates scientific and technological development (what is sometimes called in the existential risk community, “PASTA,” a “Process for Automating Scientific and Technological Advancement”), the invention of black-ball technologies may grow increasingly likely.

Estimates of existential risk are based not only on assessing the destructive capacities of various technologies but also, crucially, on assessing our ability to prevent catastrophic use of those technologies. Ord’s alarming 1/6 estimate already factors in that we would cut existential risk by half, because we would, in the coming decades, “get our act together and start taking these risks very seriously.” Any such

---

40 See generally, Iason Gabriel, Artificial Intelligence, Values, and Alignment, 30.3 MINDS AND MACHINES 411 (2020); Nick Bostrom, SUPERINTELLIGENCE (2014).
42 Id.
43 Id. at 455-56 (discussing the possibility of apocalyptic technology created “with a piece of glass, a metal object, and a batter arranged in a particular configuration”).
45 ORD, supra note 1.
awakening in our political and legal systems is inhibited by a wide range of cognitive biases that make it difficult to recognize the scope of existential risk. To cite just a few key examples, people generally find it difficult to imagine events on a scale we have never seen before (the availability heuristic) and with a moral lens that is evolutionarily tuned to small-scale and nearby harms (scope neglect) that befall known individuals (the identifiable victim effect) who are alive today (present bias).46 We tend to dismiss threats of low probability events unless our emotions are primed.47 In popular understandings of existential risk, these cognitive biases may be exacerbated by the association of existential risk with science fiction, irrational doomsayers, or, in some emerging caricatures of the current existential risk community, hyper-rational “tech bros.”48 In the political realm, these cognitive biases may be exacerbated by short-term incentives that prioritize currently living (and voting and lobbying) people over future generations.49 Moreover, existential risk


47 Cass Sunstein, Worst Case Scenarios (describing the general lack of emotional response around threats that are rare or unprecedented, and our tendency to over-react when low-probability events are emotionally salient, such as when such events have occurred recently); Cass Sunstein, Averting Catastrophe (observing that some catastrophes are the result of exponential rather than linear growth, which leads to under-reaction due to “exponential growth neglect”); Eliezer Yudkowsky, Cognitive Biases Potentially Affecting Judgement of Global Risks, in Global Catastrophic Risks 91, 91–115 (Nick Bostrom & Milan M. Cirkovic, eds., 2d ed. 2020) (summarizing the relevant cognitive biases and noting that humanity tends to “overestimate the predictability of the past and underestimate the surprise of the future”).

48 See Joshua Schuster and Derek Woods, Calamity Theory: Three Critiques of Existential Risk, https://manifold.umn.edu/read/calamity-theory/section/3a175630-820b-4886-aadc-06f3b04021516 (expressing a general wariness in relation to the observation that “existential risk theory has been conducive to a warm reception by a ‘tech bro’ Silicon Valley audience.”); Posner, supra note (arguing that irresponsible doomsday predictions can lead to a backlash of excessive optimism regarding the risk of large-scale catastrophe, but noting that thoughtful science fiction can also be helpful to illuminate these threats).

49 Tyler M. John, Representing Future Generations, YouTube (Mar. 21, 2020), https://youtu.be/095kFEA-jpE (observing that elected officials and other political leaders tend to consider future effects only on the scale of 2-5 years or “the next decade,” due to cognitive biases, time preference, and election incentives).
raises the collective action problems associated with goods that are public, global, and intergenerational (thus, the condition of “existential safety” is analyzed in this literature as an “intergenerational global public good”). These political conditions could be made far more perilous by “existential risk factors” in the coming years, such as great power wars, extreme environmental impacts, or other events that make it less likely that we, as a global community, will be willing and able to work together to address threats on the existential scale.

But maybe these “biases” against concerning ourselves with existential risk are actually pointing toward true. Should we care about preventing existential risk? Does existential risk matter? This question has both empirical and normative dimensions. This section has so far focused on empirics—assessing the probability that an existential event will occur. If this probability is extremely low or extremely high (and intractable), we might have little reason to address existential threats. The normative dimension asks whether an existential event would be undesirable. The simplest response from participants in this study is that existential risks threaten to harm a great number of people who actually exist. This includes people whose lives would be ended by an existential catastrophe or whose lives would be made much worse under an unrecoverable dystopia scenario. Moreover, efforts to mitigate existential risk will often overlap with efforts to mitigate sub-existential threats, which bolsters the case that working on existential risk matters to people who actually exist. The more complex normative response considers the impact of human extinction on future generations, where the impact would be “felt” by people who would not have the opportunity to exist. A full treatment of this issue and related theories of population ethics is beyond the scope of this article. This issue is the subject of a new field of philosophical inquiry known as “longtermism,” which considers the moral weight of future generations. But it is worth briefly noting that many participants in this study analyze this question in the following terms: potential future persons could outnumber us (current persons) to such a radical degree that if one assigns any non-negligible value to the existence of future persons, and if one believes that future persons will tend to have lives worth living, one might

50 See Ord, supra note 1.

51 See generally Derek Parfit, Reasons and Persons (2d ed., 1986) (introducing population ethics with the hypothetical comparison of an event that ends 100% of human life and an event that ends 99% of human life, where the latter event permits the continuation of future generations, thus raising the question of how much we should value people who could one day exist).

view human extinction as an enormous harm.\(^\text{53}\) Combining these concerns relating to human extinction scenarios (for the actual people who would die in such a catastrophe and the potential future people who would not have a chance to exist) and permanent dystopia scenarios (for many actual people current and future) makes a strong case, as participants in this study argue, that existential risk matters and should be one of our leading global priorities.

### B. Socio-legal literature on law and social change

The findings in this article will describe the model of social-change lawyering being developed within the movement to mitigate existential risk. As context for this discussion, this section provides a brief background on how socio-legal scholars have described the role of lawyers in empirical studies of other social movements.\(^\text{54}\) Much of this literature has criticized lawyers for exaggerating the value of court-led social change, especially where lawyers fall under a “myth of rights” and a “hollow hope” that de jure victories in the courts will, on their own, bring about de facto social change.\(^\text{55}\) Even some of the most celebrated lawyers of the Civil Rights Movement are cited as examples of this overly legalistic orientation, prioritizing litigation while discouraging other tactics (e.g. grassroots organizing and legislative advocacy).\(^\text{56}\) These scholars cite examples where victories in the courts have sparked backlash and countermobilization, undermining any lasting social change.\(^\text{57}\) Moreover, some scholars note that lawyers tend to dominate movement agendas, marginalizing the

---


57 See, Rosenberg, *supra* note.
grassroots voices of the most affected constituencies. 58 These observations have led to calls for lawyers to take a reduced role in movement leadership. 59

In contrast, recent empirical accounts challenge the rendering of social-change lawyers as narrowly legalistic and strategically unsophisticated. 60 Scholars increasingly see a “revival” of movement lawyering under the rubric of “integrated advocacy,” wherein lawyers coordinate their distinctively legal work (which includes not only litigation but a wide range of services that draw on legal expertise) with contributions to other strategies such as movement building, shaping public opinion, and advancing new legislation. 61 Professor Scott Cummings has described in empirical detail several examples of this “new canon” of social-change lawyering. 62 He notes that the marriage equality movement in particular has reshaped the test case litigation model with a greater emphasis on fostering favorable public opinion (“hearts and minds”) through local legislative campaigns. 63 This creates a sense of collective demand for reform, which may help to persuade judges to make bold decisions (e.g., Obergefell v. Hodges 2015) while also persuading the public to support enforcement of those decisions with relatively little backlash. 64 Cummings stresses that most movements are long-term political projects, marked by continual dynamics of resistance and struggle. 65 Law can play an effective role in these ongoing social-change projects, especially when legal efforts are coordinated within an integrated framework of other strategies. Moreover, by embedding lawyers within movements

58 See Bell, supra note.
59 SCOTT CUMMINGS, MOVEMENT LAWYERING forthcoming (describing the “movement liberalism” literature where scholars recommend that social-change lawyers take a more limited, conventional client-centered advocacy role in support of grassroots organizations).
60 See Alan K. Chen, Rights Lawyer Essentialism and the Next Generation of Rights Critics, 111 Mich. L. Rev. 903, 905-06 (2013) (describing “rights lawyer essentialism” as a common but inaccurate portrayal of civil rights attorneys as “elitist, singularly minded litigation hawks who care little for their clients or the subtleties of the dialectic political process.”); see also ALAN K. CHEN AND SCOTT L. CUMMINGS, PUBLIC INTEREST LAWYERING: A CONTEMPORARY PERSPECTIVE 518 (2013) (providing examples of how cause lawyers conceive of litigation “not in isolation, but as part of a comprehensive set of tools that are useful in advancing social reform.”).
61 SCOTT CUMMINGS, LAWYERS AND MOVEMENTS (forthcoming) (observing that movement lawyering is experiencing a “revival” after “decades of dormancy”).
62 Id.
63 Id.
64 Id. (observing that lawyers can and should contribute to all aspects of this integrated model, and can justifiably take leadership positions where they should be subject to the same scrutiny as other movement leaders).
65 Id.
and larger strategic conversations, integrated advocacy serves to enhance lawyers’ accountability to the populations most affected by an issue.

My effort to bring this literature to bear on the legal activists examined in this article may raise an initial objection that I want address here at the outset. This literature has focused on lawyers within social movements, especially progressive grassroots movements where aggrieved populations find a collective voice and seek remedies often through disruptive tactics and putting pressure on power holders.66 But the community working on existential risk has not (at least not yet) developed into a broad social movement. As emphasized by participants in this study, the population most affected by existential risk is future generations, who cannot form a grassroots movement on their own behalf.67 But, in some respects, this voicelessness resembles the classic civil rights concern for populations that lack access to legal and political power. Future persons cannot vote, bring suit, hold public demonstrations, or bargain to advance their interests.68 Some participants in this study have backgrounds in progressive social justice movements, and see their work on existential risk as an effort to counteract discrimination against future persons—in recognition of the “tyranny of the present over the future”69 Others frame the issue in terms of equality, as one participant explained: “Our assumptions are quite simple. We want to treat everyone as equal, not just in space, but also in time.” This suggests some commonality with the progressive social movements examined in socio-legal literature—including those movements that have been


67 Toby Ord, Senior Rsch. Fellow, Oxford Univ & Author, The Precipice, Stanford Existential Risks Initiative Virtual Conference (Apr. 17, 2021) (noting that the population who “bear the most of the relevant costs” would usually vocalize their interest and “campaign and push for change.”).

68 MACASKILL, supra (observing that future generations “can’t bargain or trade with us, so they have little representation in the market.”).

69 Nick Bostrom, Existential Risk Prevention as Global Priority, 4 GLOB. POL’Y 15 (2013), https://doi.org/10.1111/1758-5899.12002; ORD, supra note 1 (arguing that our neglect of future generations “privileges a tiny minority of humans over the overwhelming majority yet to be born”); SUNSTEIN, WORST CASE SCENARIOS, supra note (proposing a Principle of Intergenerational Neutrality, requiring “members of one generation to give equal weight to the interests of those who will follow”).
studied with roughly the same qualitative research design employed in this article.\(^7^0\)

Moreover, participants in this study, as represented in the findings below, are engaged with the same strategic questions debated throughout this literature, e.g. how much to invest in broader movement building, how to address reputational risks and the threat of backlash, how to enhance accountability to key constituencies, and how important it would be to secure justiciable rights for the population of concern (future generations) as opposed to addressing more specific issues. With these commonalities in mind, this article will refer to the community working on existential risk as a nascent “movement.” This allows for an inquiry into whether, and to what extent, insights from socio-legal literature apply to the novel context of existential risk mitigation. In particular, this study contributes to the literature on early-stage movements and “legal mobilization,” which analyzes the question of why, when, and how movements decide to deploy law.\(^7^1\)

II. RESEARCH DESIGN

Ethnography was conducted at the Legal Priorities Project (“LPP”) over a period of five months in 2021 and 2022, during which 59 virtual events were observed. Ethnography is an anthropological method of “participant observation,” in which the researcher is invited to join a community while taking detailed analytic and descriptive notes on norms, interactions, and other cultural dynamics, as well as the researcher’s own experiences of membership in the community.\(^7^2\) Rather than testing hypotheses, ethnographers tend to inductively generate “grounded theory,” wherein the study’s key theoretical observations emerge from the thematic coding and analysis of fieldnotes.\(^7^3\) This requires deep immersion in the culture under study. LPP was highly receptive to this methodology. They permitted me to observe weekly all-hands and small-group meetings, research workshops, and textual discussions and shared documents, including their online messaging platform and a range of internal reports. I also regularly engaged, in a non-ethnographic capacity,


\(^7^1\) See generally, Emilio Lehiucq & Whitney K. Taylor, Conceptualizing Legal Mobilization: How Should We Understand the Deployment of Legal Strategies?, 45 LAW & SOC. INQUIRY 166, 168 (2020).

\(^7^2\) See KAREN O’REILLY, ETHNOGRAPHIC METHODS (2012) (suggesting that ethnographers combine “emic” understandings, from the perspective of the subjects, with the researcher’s own “etic” understandings rooted in their research questions and interests).

\(^7^3\) See JULIANNE S. OKTAY, GROUNDED THEORY (2012).
with LPP and the larger legal community working on existential risk for a one-year period in preparation for this article. This included visits to key sites of the movement (Oxford, London, Geneva, and Cambridge, Massachusetts) and other in-person meetings. I attended twelve conferences and workshops over this period, while regularly following online discussions (e.g., Slack workspaces, Facebook groups, blogs, Discord channels, podcasts, newsletters, and the Effective Altruism Forum). A range of published online materials are also referenced throughout this article, including white papers, research agendas, and curricula for courses and reading groups.

Over the course of this year of data collection, I also conducted 53 semi-structured interviews. The interview participants included LPP team members, affiliates, and summer research fellows, as well legal and political advocates at other organizations working on existential risk. Some interviews were held in-person but most were remote (via video-conferencing calls) with participants who were physically located in all continents of the globe except Antarctica. The interview protocol began by asking participants for a biographical account of how they became interested in the topic of existential risk and related concerns for future generations. This was followed by questions about how they perceive existential risks, the community working in this area, and the organizational cultures they have encountered, including considerations of epistemics, dissent, emotions, and identity (as these cultural traits are categorized infra Part V). The interviews then transitioned to a discussion of legal strategy. Participants who are lawyers were also asked about how their professional identities comport with their identities as activists for the mitigation of existential risk. In addition to these formal interviews, I engaged in hundreds of hours of informal conversations with members of this community, which are not quoted or paraphrased here but nevertheless inform the empirical analysis.

With qualitative methods, the researcher serves as the research instrument in the field. Rather than, for example, distributing a survey, the researcher directly asks participants their questions and observes participants in their settings, filtering the entire data collection process through the researcher’s own perceptions and biases. This fact, along with a long line of critical literature about the power dynamics of representing research subjects, particularly considering the historical association between ethnography and colonialism, has led to a call for qualitative scholarship to be accompanied by a reflexive account of the researcher’s interests, goals, and positionality. In this spirit, I will briefly provide relevant observations here—more auto-ethnographic notes can be found throughout the presentation of findings infra.

74 This statement assumes the convention of categorizing seven continents as North America, South America, Africa, Australia, Asia, Europe, and Antarctica.
Over the course of the study, I transitioned from an outside observer to a more active participant in strategic discussions in this field. Following the conclusion of the ethnography, I have continued to volunteer with LPP, contributing to a range of the organization’s strategic discussions. I generally agree with the notion that existential risk matters and that this community is taking useful steps toward an effective response (see recommendations infra Part VI). Becoming a “member” is not uncommon in ethnography, nor is it necessarily discouraged, although it can predispose the researcher to portray the community under study in a more favorable light. This tendency is somewhat diminished by LPP’s strong cultural commitment to dissent and criticism (see infra Part V). In the LPP meeting in which I formally requested permission to conduct ethnography, I expressed my intention to contribute to socio-legal theory and to provide feedback to the organization and the larger movement around existential risk. I explained that the project would draw comparisons to how law has been deployed in other movements for large-scale change, including movements for civil rights and animal protection that I have studied in other research. LPP members embraced this project from the start and, to their credit, encouraged me to provide candid and critical feedback. I presented preliminary findings to the LPP team several times and invited participants to comment on an earlier draft of this article. This collaborative approach, in combination with assurances of confidentiality, helps to promote trust and transparency in the interaction between researcher and participants. My relationship with this community has been consistently warm and mutually respectful. This sense of rapport likely colors some of my interpretations, although my ultimate sense of responsibility as an empirical researcher is to provide an accurate description of this community, including its shortcomings and points of tension.

75 See Hsun-Yu Sharon Chuang, Complete-Member Ethnography: Epistemological Intimacy, Complete Membership, and Potentials in Critical Communication Research, 14 INT’L J. QUALITATIVE METHODS, Nov. 18, 2015, at 1, 2 (discussing the empirical value of becoming a “member” in ethnographic studies); Cf., Margaret D. LeCompte & Judith Preissle Goetz, Problems of Reliability and Validity in Ethnographic Research, 52 REV. EDUC. RSCH. 31, 46 (1982) (detailing the empirical and ethical risks when researchers develop strong social relationships with participants).

76 Identifying information is used in this article only where specific permission was granted by the identified participant.
III. THE PRIORITIST THEORY OF EFFICACY

Several months into the qualitative data collection for this project, I started using the label “prioritist advocacy” in my field notes to describe the model of social-change lawyering that I was observing. This model is “prioritist” in the sense that it relies on a methodology for identifying the highest priority goals that could be pursued (e.g. the mitigation of existential risk) and the strategies that maximize impact toward those prioritized goals. I considered alternative labels to describe this form of advocacy, such as “maximizing” “altruistic,” and “epistemic,” which speak to foundational concepts described below. But none of these alternative labels captured this movement’s unique focus on rigorous prioritization processes at every level of decision-making. Their formula for setting priorities is an innovative development within the tradition of social-change lawyering. As part of my collaborative research method, I asked participants for feedback on this question of how to label their theory of advocacy. This was discussed in a dedicated all-hands discussion at an internal meeting of the Legal Priorities Project (“LPP”). Most participants agreed with my choice of “prioritist advocacy.” A German participant jokingly offered a single-word translation of the concept: “Interessenvertretungspriorisierung.” This Part defines the methodology of interessenvertretungspriorisierung, although for the sake of brevity I will continue to use the English term, “prioritist advocacy.” This model of movement lawyering can be broken down into a two-step process of selecting goals (reviewed in Section A) and then reverse engineering strategies from those goals through a counterfactual analysis (reviewed in Section B).

A. Selecting Goals

How do social-change lawyers decide what issue they will address? In most cases, the answer from socio-legal scholarship is relatively intuitive. Lawyers, like other activists, are attracted to causes because of some combination of identity, values, ideology, and the availability of resources and opportunities. Generally, there is already a demand from communities seeking remedy or reform. Lawyers then work to translate this demand into legal action. This approach can produce powerful results, as reflected in a long line of successful campaigns for civil rights and other causes. But it can also lead to well-intentioned yet ineffective efforts, including where lawyers fail to consider alternative cause areas that are larger scale or more amenable to impact. Scholars of movement lawyering have called for more formal

---

77 This participant noted that Google Translate “has a good sense of the word” with its translation back to English as “advocacy prioritization.”

78 See generally, STUART A. SCHEINGOLD & AUSTIN SARAT, SOMETHING TO BELIEVE IN: POLITICS, PROFESSIONALISM, AND CAUSE LAWYERING (2004).
processes for “identifying objectives and establishing priorities among them.” The prioritist advocacy model provides one answer to this call. The prioritists take the unusual starting place of “cause neutrality,” meaning that they could work on any conceivable issue or cause. They then undergo a systematic process for selecting the cause area where they predict that they can have the greatest moral impact. As discussed below, this process is an application of the “priorities research” methodology of “effective altruism,” a philosophical and social movement that seeks to maximize the good that can be accomplished through philanthropy, careers, advocacy, and other means. The existential risk community, as currently constituted, can be viewed as a subset of effective altruism. The participants in this study are the first legal movement to apply effective altruism to the law and to legal activism.

Following the cause prioritization methodology of effective altruism, the lawyers in this study begin with first-principles of morality, which is most often framed in this community as a mandate to maximize well-being, although they draw on other philosophical foundations as well. Next, the aim is to determine what cause areas

79 Cummings and Rhode, Public Interest Litigation: Insights from Theory and Practice (observing that public-interest law generally lacks formal processes for prioritizing objectives and that “few organizations operate with explicitly articulated theories of change or specific measures of performance”).

80 See generally MACASKILL, DOING GOOD BETTER at 14; Benjamin Todd, Can One Person Change the World? What the Evidence Says, 80,000 Hours, https://80000hours.org/career-guide/can-one-person-make-a-difference (Apr. 2017) (“People often wonder how they can make a difference. . . [but] the key question is, ‘how can I make the most difference?’”). Effective altruism has grown to include a number of non-profit organizations focused on research (e.g. the Global Priorities Institute; Rethink Priorities), career advising (e.g., 80,000 Hours, Training for Good), assessing effective philanthropy (e.g. GiveWell), grantmaking (e.g. Open Philanthropy, Longview Philanthropy, Future Fund, BERI), and education and outreach (e.g. the Centre for Effective Altruism, which hosts several regional and global conferences every year, runs the Effective Altruism Forum, and supports other effective altruist organizations).

81 That existential risk is primarily being addressed through an effective altruist lens is partly explained by historical happenstance—the academic origins of effective altruism and existential risk studies have coincided in just the past couple decades through an overlapping group of founding figures, many of whom are associated with Oxford University where they have established institutes and non-profit organizations. Founding Oxford faculty in these overlapping fields include Nick Bostrom, Hilary Greaves, Will MacAskill, and Toby Ord. Organizations established around Oxford include the Centre for Effective Altruism, the Global Priorities Institute, and the Future of Humanity Institute, all of which currently share an office space.

82 See, e.g. LPP research agenda (citing grounds for protecting future generations primarily in utilitarianism but also in deontology, virtue ethics, and other traditions of moral and political philosophy).
are amenable to the greatest impact toward their moral goals. The core criteria of this methodology are importance, neglect, and tractability (the “INT” analysis). In this framework, a cause is prioritized not only because it affects many people and to a great degree (importance), but also because it is feasible to reduce this risk without imposing morally offsetting costs (tractability), and because the issue is not already receiving adequate attention such that any interventions would be subject to diminishing returns (neglect). For example, in recent discussions that I observed within the impact litigation team at LPP, engineered pandemics are favored by this INT formula because such pandemics could produce widespread illness and death affecting current and future generations (importance), there is promising evidence that litigation could be useful in addressing issues such as unsafe laboratory practices (tractability), and the governmental budgets and overall public attention to the issue are greatly lacking (neglect). For another example, this same impact litigation team has tended to view existential threats from AI as highly important and neglected but less tractable, because we do not yet have a clear enough sense of how to create safe AI, nor how litigation can contribute to this cause. This analysis could change if the AI safety research community provides more specific recommendations for law and policy.

Existential risk has recently come to be identified as a top collection of cause areas in the effective altruism community, largely owing to the importance of the issue for a great number of current lives and a far greater number of future lives (but also because of an analysis of neglect and tractability). This concern for future generations follows from a commitment to “moral circle expansion.” In the early years of effective altruism (the 2000s and early 2010s), the community was focused on expanding the moral circle across spatial boundaries, leading to a greater concern with poverty and suffering beyond local and domestic contexts. Thus, effective altruism has sought to direct philanthropic dollars to the most effective global poverty and health interventions. Empirical evidence suggests that it currently costs roughly two to five thousand dollars, on average, to “save a life,” which can be defined in different ways but in practical terms can be seen in the example of stopping a child from dying of a preventable illness where the child would then lead a full

---

83 See MACASKILL DOING GOOD BETTER, supra note (outlining the “INT” analysis and proposing five key questions to assess these criteria: “How many people benefit, and by how much? Is this the most effective thing you can do? Is this area neglected? What would have happened otherwise? What are the chances of success, and how good would success be?”).

84 See MACASKILL, DOING GOOD BETTER, supra note (explaining the neglect criterion by reference to the diminishing returns of investing in issues that are already receiving a great deal of attention, the potential for counterfactual impact when otherwise few resources would be spent on an issue, and the observation that the most impactful cause areas are often overlooked).
healthy lifespan. More recently, the community has applied moral circle expansion across species boundaries in recognition that non-human animals, particularly in agricultural contexts, are suffering in a morally relevant and large-scale manner. And in just the past few years, this community has taken moral circle expansion across temporal boundaries, with a focus on future generations. Although it is difficult to know how most of our actions affect future generations, existential risk is arguably less susceptible to this concern—foreclosing the future entirely or locking us into an unrecoverable dystopia are permanent effects extending into the long-term future. Although existential risk is generally favored among effective altruists at the moment, the prioritist methodology can, by design, yield different answers. One of the core commitments of this framework is a willingness to update and change directions upon new information.\(^{85}\) One participant explained: “If you were to prove that one of the core [effective altruist] cause areas actually wasn’t an issue, like that AI safety was not a big risk, you would be celebrated.”\(^{86}\)

The participants in this study are developing a distinctly legal variety of this cause prioritization methodology, what LPP refers to as “legal priorities research,” in recognition that the tractability and neglect analyses may give different answers when considering what can be accomplished with legal interventions.\(^{87}\) For example, some participants noted that existential risk is currently receiving greater attention in public policy, with new legislation and inter-governmental talks around protecting future generations, but is neglected in courts.\(^{88}\) If this neglect is combined with a determination that legal advocacy can be impactful (tractability), this formula would recommend expanding legal efforts to mitigate existential risk.\(^{89}\) The question of how much to prioritize such legal efforts is still the subject of a great deal of research and

---

85 Winter et al., supra note (“we offer a rigorous yet flexible, and potentially ever-evolving methodological framework for deciding which problems to work on and how to tackle them.").

86 MACASKILL, supra note (“we genuinely just want to do what’s best for the world, so if we’re wrong about anything — even if it’s the thing we’ve been dedicating our lives to — we should want to know.").

87 Winter et al., supra note (observing that “some cause areas may prove to be significantly more neglected or tractable from a legal perspective and thus become a top priority for law.").

88 As one participant noted: “For all the people doing policy outreach focus to legislatures, no one focuses on judges...” Another participant noted that existential risk is “reliably neglected in judicial decision-making.”

89 In addition to this observation that neglect motivates LPP’s plans for legal interventions, it is also noteworthy that neglect is a major factor in LPP’s decisions about where to focus their research efforts (citing lack of attention from legal scholars), educational programming (citing a lack of legal curriculum), and public opinion (citing the lack of public awareness and discussion).
debate among the participants in this study. But just by investing so heavily in this process of selecting the right cause area, prioritist advocacy represents a departure and innovation within the tradition of social-change lawyering. I will return to these broader implications of the model in the discussion section (infra Part VI).

B. Selecting Strategies

Having selected a cause area where the prioritists predict that they can have the greatest impact (e.g. existential risk), the next step is to select strategies to optimally advance the prioritized cause. This step is visually represented in Figure 2. As explored in this section, some aspects of Figure 2 may appear familiar from the leading recommendations of the literature on law and social change. This literature advises lawyers to think of their legal tools as just one part of a larger campaign for social change, where the ultimate focus is on achieving lasting, de facto impact. Prioritist advocacy can be viewed as an effort to maximize in the direction of this recommendation from the literature.

Figure 2. The Prioritist Framework for Selecting Strategies

As evident in qualitative observations, the prioritists begin with a process that I label “end goal primacy.” This involves working backward from their cause (the “end goal”) to more specific and immediate goals that are preconditions to achieving the end goal. This approach is derived from “Theory of Change,” a framework common

---

90 LPP’s website describes their approach as “mission oriented,” whereby they “work backwards from these long-term goals to prioritize the projects that we think are most
in philanthropy and impact investing, which recommends visually mapping out the reverse engineering process. In an internal document, LPP has created such a map with an end goal to advance a state of the world in which “humanity’s long-term potential is safeguarded.” The report notes that this goal is “far too vague to guide our decision-making” but that it helps the organization “link every subsequent step back to our longtermist roots.” To take the example of just one causal chain through this document, they list “longtermist research is known and valued by policy-makers” as one precondition to achieving their end goal. This is preceded by building relationships with policy makers and encouraging LPP affiliates to pursue political careers, which is preceded by building relationships with organizations focused on policy. The key point here is that more specific and immediate goals are defined after higher goals. If new information alters higher goals, this model requires a willingness to reconsider an organization’s immediate and day-to-day strategic priorities.

When designing strategies, the prioritists draw from key insights in the “integrated advocacy” literature. Scholars in this socio-legal literature advise lawyers to blur the distinction between law and policy and to coordinate strategies and frames across the law/policy divide. For the prioritists, this integration is consistent with the effective altruist notion of a “portfolio approach” and an “alliance mentality,” emphasizing the overall impact of collective efforts toward a particular end goal, rather than the isolated impact of the actions taken by any particular promising and impactful.” https://www.legalpriorities.org/opportunities.html (last visited January 11, 2022).

91 See Edward T. Jackson, Interrogating the Theory of Change: Evaluating Impact Investing Where it Matters Most, 3 J. SUSTAINABLE FIN. & INV. 95, 100 (2012) (recommending visual mapping to identify “underlying logic, assumptions, influences, causal linkages, and expected outcomes of a development program or project”).

92 Other organizations working on the long-term future of humanity have also applied Theory of Change. For example, the Simon Institute for Longterm Governance published a diagram that lists an end goal of “long-term human flourishing” with preconditions to improve “long-term institutional fit” through improving decision-making process and integrating longtermist concerns in “dominant societal narratives,” “institutions,” and “policy agendas.” https://www.simoninstitute.ch/blog/post/our-theory-of-change/ (last visited August 12, 2022).

93 See Cummings, supra note (observing that the “new convention” among public interest lawyers is to engage in “multidimensional” and “integrated” advocacy where lawyers are “strategically sophisticated” and work with a wide range of allies to “advance political goals in multiple venues through coordinated tactics in the face of persistent opposition”).

94 See Cummings, supra note (reviewing the interdisciplinary literature on the “interaction between law and politics” and recommending that law be “coordinated with politics through an integrated strategy that maximizes the potential for sustainable social change.”).
individual or organization. An internal LPP report on impact litigation directly references the integrated advocacy literature and emphasizes that litigation efforts should be paired with complementary actions in the domains of legislation, policy, public opinion, academic research, and education.

For socio-legal scholars, the blurring of law and policy is meant as a corrective to the long-standing concern that lawyers in social movements tend to over-emphasize the value of law and courts, falling under a “myth of rights” while discouraging legislative outreach, grassroots organizing, and other forms of activism. But the myth of rights does not seem to hold much appeal among the prioritists. Some participants in this study noted that, until very recently, lawyers in effective altruism were more drawn to the opposite myth: that, for effective altruist causes, as one participant put it, “litigation is useless, law is useless.” Until recently, law students who were committed to effective altruism were primarily advised to take high-paying positions so that they could “earn to give” to effective charities rather than seeking out impactful legal work relating to prioritized cause areas. But this trend has changed over the past few years. The effective altruism community now appears to see much more value in legal and political efforts. Some hesitations about the role of law remain, although the legal efforts are receiving a great deal of funding and support. These lingering hesitations may help this community avoid falling into a myth of rights. One of the pathways to impact being considered by participants in this study is to seek to create justiciable rights for future generations in domestic constitutions, common law doctrines, international law (e.g., extensions of human rights across time), and inter-governmental agreements. But participants were consistently mindful that simply establishing such rights would not necessarily mean they would be enforced in transformative or otherwise meaningful ways.

A central commitment of the prioritist approach to assessing strategies is that they undertake counterfactual analysis, whereby they assess the marginal expected

---

95 See MacAskill, supra note (“The fact that we each act as part of a wider community warrants a ‘portfolio approach’ to doing good—taking the perspective of how the community as a whole can maximize its impact.”).

96 See Scheingold and Sarat, supra (noting that social-change lawyers tend to be “attracted to courts as fly paper,” mesmerized by a “myth of rights”).

97 Participants working in legislative outreach made very similar observations about effective altruists’ historical lack of interest in the policy domain. For example, one such participant observed that when they attended effective altruism conferences just a few years ago they would rarely find anyone with an “interest in living in DC or working with the US government.”
value of any action under consideration.\textsuperscript{98} This involves weighing how one option compares to alternatives, including the possibility of doing nothing (“no action” in Figure 1 above).\textsuperscript{99} This approach draws from the norms of effective altruism, where proposals are generally accompanied by an analysis of “what would have happened otherwise.” In my observations at LPP, this counterfactual approach was consistently evident in debates over planned legal interventions. As described in an LPP report, any strategic action should be compared to “feasible alternatives and counterfactuals,” including consideration of “the chances the legislature or executive will take up this issue at some point anyway” or whether these governmental entities will act “with enough haste.”\textsuperscript{100} For example, when considering whether to provide formal comments on the UN Declaration on Future Generations, LPP engaged in a lengthy decision-making process about whether such a commentary would be counterfactually impactful when considering what would happen otherwise. What is expected impact of the Declaration? Would LPP’s comments be expected to make the Declaration more impactful? Would the “people hours” spent on this project be more valuable than spending those hours on other projects? The LPP team debated these questions at length during a four-day in-person “theory of change retreat” that I observed.

When comparing strategies to mitigate existential risk, the prioritists give considerable weight to whether any particular strategy helps to overcome cognitive biases. This is evident in discussions of specific strategies as well as high-level discussions about whether to focus more on law (e.g., litigation and judicial outreach) or policy (e.g. legislative and regulatory advocacy). A participant working in the policy space emphasized that their work is often stymied by the range of cognitive biases that inhibit recognition of existential risk, noting that “99.5% of policy makers don’t actually realistically think in the next 100 years human beings could possibly go extinct....” Another participant working in this space emphasized that elected officials are focused on the near-term demands of their constituents and, more generally, have “no time” because they have people “hounding them...every second of the day” and an incessant “huge box of materials to go through.” Thus, when presenting existential risk to lawmakers, one participant reported commonly receiving the response: “How can I think about this when we’ve got so many

\textsuperscript{98} See William MacAskill, Doing Good Better 57 (2015) (describing “marginal utility” as a “fundamental piece of scientific reasoning” that motivates the methodology of effective altruism).

\textsuperscript{99} As one participant explained, legal actions should be evaluated according to the extent to which they have an “independent or a counterfactual and positive impact.”

\textsuperscript{100} LPP internal litigation report, on file with author.
crocodiles closer to the boat?" Participants working on the legal dimension of these issues cited further law-specific biases, including the possibility that law has a conservative, backward-looking precedential orientation that might be unreceptive to the “new legal techniques” that may be required to address novel risks to future generations. Moreover, participants worried that the judiciary might struggle with the probabilistic nature of existential risk given their lack of “formal training in quantitative subjects.” But participants also saw promise in the judiciary to care for future generations as a disenfranchised group. As one participant noted, courts uphold the “liberal values” of democracies and the protection of the “political minority.” Several participants expressed optimism that judges, owing to their general reliance on “abstract values,” might be receptive to expanding human rights “independently of time.” It is also important to note that the prioritists are taking actions that directly seek to reduce these cognitive biases, including holding educational workshops for policymakers (and planning similar workshops for judges) on existential threats and the distinct decision-making challenges around low-probability/high-impact risks.

So far, this section has described the ex ante phase of the prioritist method—predicting how their planned actions might optimally advance an end goal. The prioritist effort to maximize impact also demands continual reassessment of actions as they are being pursued. This involves Bayesian updating and the development of metrics and even experimental methods when possible. This emphasis on

---

101 As one participant working in policy outreach explained: “Most people in political realm are working on 24-hour review cycle. There’s a lot to be done in the short-term, so its hard to get anything done in the long-term. Long-term is a year out from now. That’s how chronologically structured political folks think about things. But with respect to 100 years that would be a really hard case to make.”

102 See Albiston, supra (observing that lawyers often “deradicalize and subtly reshape social movements”); Cummings, supra note (noting that legal framings can “sanitize” issues to “comport with mainstream values,” perhaps because, as critical legal scholars have long argued, law favors the status quo and legal victories tend not to transform structural relations); Rhode and Cummings, supra (noting that legal actions can dissipate grassroots activism, thereby reducing a “movement’s transformative potential”).

103 Witner et al, supra note (providing examples of non-quantitative reasoning in the law, including standards of “beyond a reasonable doubt,” “probable cause,” and “balancing tests.”)

104 An internal LPP report makes that case that “specific measures of performance” are necessary to assess “progress…made towards our goals in an observable, measurable way” and to “identify where theory and practice fall apart and thus where we should rework our [Theory of Change]. Experimental projects in this field include a forthcoming study of how judges may interpret legal arguments relating to existential risk cause areas, manipulating a number of factors including source and content of law, level of
empiricism is a distinct feature of the prioritist model. But the prioritists also draw on some more familiar considerations from socio-legal literature, which emphasizes the different ways of conceiving of impact. As applied by participants in this study, this includes questions of whether to pursue broad and holistic approaches to existential risks (e.g. establishing rights and legal standing for future generations, criminalizing activities that generate existential threats, influencing regulatory policy that affects multiple agencies and departments) or more narrow and cause-specific legal challenges (e.g., enforcing regulations on a company or laboratory developing a dangerous technology in an unsafe manner). They also follow the socio-legal distinction between direct effects (e.g., how a court order is enforced) and indirect effects (e.g., how a court order reshapes political discourse). As described in LPP internal documents, some legal actions may be especially valuable for their indirect effects, for example where a litigation victory or loss serves an educational function (e.g., to reveal problems and “dramatize abuses”), a motivational function (e.g., what socio-legal scholars call “internal effects” where a legal victory helps stimulate growth in the movement), a symbolic function (e.g. providing recognition of the importance of certain issues or arguments, legitimating ideas in the movement, and gaining control over issue definition), and a resource mobilization function (e.g., attracting funding and increasing bargaining power). These indirect effects may be particularly important in the context of international “soft law,” which is considered by participants a key avenue for addressing the global dimension of existential risks. As one participant noted, even if soft law does not have the teeth of strong enforcement mechanisms, it can help “inform the common language that people use...how they perceive the future...and whether or not they consider extreme risks.”

One of the challenges in evaluating efforts to protect future generations is that any action taken today may tend to “wash out” over time or even exacerbate threats in ways that are only visible in the future. In other words, assuming that humanity survives for some time, we may want a long-term plan to mitigate existential threats, but such a plan is difficult to design based on only current information. Participants emphasized the importance of creating adaptable legal and political mechanisms, described by one participant as “evergreen policies...[that are] able to

expert disagreement, how broad the rule is, and the level of abstraction at which a choice is presented.

105 Participant acknowledged the weakness of international soft law. For example, one participated cited the failures of international health regulations, which they described as “the most advanced form of regulation in the international system” and as “theoretically binding,” but noted that these regulations were “not respected during Covid.”

106 See Greaves and MacAskill, supra note.
adjust...as the technology changes.” LPP members noted that some legal interventions may persist into the distant future because of the “path dependence of law,” including certain legal devices that are explicitly “eternal.” But these same participants consistently acknowledged that any long-lasting law or policy could have the downside of limiting option value for future generations, who will have more information about their own values and preferences than we can know today. As one participant explained, we are not entirely “clueless” about how to regulate existential threats nor about some of the basic needs of future generations (to live and to have lives worth living), but we should be wary of “over-confidence.” This participant went on to observe that if majoritarian values had been permanently enshrined in our legal systems several hundred years ago, the law would still permit “slavery, human sacrifice, hunting witches.” Several participants favored what longtermist theorists have called the “long reflection,” a period of deliberation over values and mechanisms of governance, which could last centuries or millennia. These complexities when considering the long-term future may help to explain the prioritist commitment to such an elaborate and rigorous process for evaluating the expected impact of any given decision.

IV. THE PRIORITIST THEORY OF ACCOUNTABILITY

The previous Part described the prioritist methodology as it relates to efficacy and the maximization of impact. This Part considers the other key criteria by which scholars have assessed social-change lawyering: accountability. Even the most celebrated civil rights advocates have been charged, in some scholarly accounts, with failing to hear or heed the voices of the populations most affected by an issue. Recent scholarship acknowledges this “elite critique” but tends to offer an increasingly collaborative portrait of activist lawyers, in many contexts, working alongside movements and seeking to secure the remedies desired by key constituencies. These questions of accountability raise a novel issue for

---

107 LPP research agenda (citing the path-dependent and conservative nature of law, and examples of long-lasting legislation, such as the German criminal code of 1871).

108 This participant acknowledged in the discussion that modern slavery persists in spite of legal prohibitions.

109 See ORD, supra note.

110 See Derrick A. Bell, Jr., Serving Two Masters: Integration Ideals and Client Interests in School Desegregation Litigation, 85 YALE L.J. 470, 489–91 (1976) (noting that the lawyers of the Civil Rights Movement chose to prioritize racial integration in schools while Black Southerners expressed a clear preference to focus on educational quality and other approaches to addressing racial subordination in schools).

111 See CUMMINGS, supra note.
participants in this study who conceive of their primary constituency, essentially their clients, as the multitudes of people who could exist in the future. Some participants described their advocacy within the traditions of civil rights, protecting future generations in an effort to, as one participant put it, “give voice to people who have been underrepresented.” But how can these advocates be accountable to a population that does not (yet) exist, and so cannot speak? Moreover, this movement’s accountability is complicated by the question of how to include the various perspectives of currently living people—a population that is also affected by existential risk and the costs of interventions. This raises the central puzzle explored in this Part: How do these advocates conceive of their accountability to current people (who are capable of speaking) and future people (who are the greater affected population but incapable of speaking)?

Accountability to clients and constituencies is a well-developed topic in the literature on the professional responsibility of lawyers. First-order representation issues, which relate to duties to clients, raise an inherent tension between the lawyers’ justified paternalism, owing to their expertise, and clients’ justified autonomy and control over objectives. In the context of social-change lawyering, these tensions can be heightened where attorneys prioritize the broader impact of a case over securing remedies for the particular client. Participants in this study often anticipated such first-order issues, noting that their litigation plans may tend

---

112 It is worth noting that many participants in this study would frame the issue as a concern not only for future humans but also for future non-human animals, possible post-humans, and possible sentient AI.

113 See also, ORD, supra note (calling for efforts to bring “the representation of future generations into national and international democratic institutions.”).

114 Note that the notion that future persons are the greater affected population implies either existential catastrophes that leave future persons to live under dystopic conditions or, under a human extinction scenario, a belief that the existence of potential future persons matters (i.e. that human extinction would cause harm beyond ending the lives of the people who would experience the extinction event). See discussion supra Part I.


to focus on relatively remote (future) impacts of a case, which may be orthogonal or opposed to the interests of an individual client.\textsuperscript{117}

But participants were more troubled by representational issues of the second order, regarding duties to constituencies, causes, and a sense of “public accountability.” These second-order duties are central to the practice and theory of “cause lawyering,” which refers to lawyers who work on behalf of causes and social movements.\textsuperscript{118} The image of the lawyer as an amoral hired-gun for clients is replaced in this tradition by the lawyer’s own “moral commitment.”\textsuperscript{119} In a wide range of movement lawyering contexts, this approach has raised difficult questions about how, and to what extent, key constituencies should be included in strategic decision-making processes.\textsuperscript{120}

The participants in this study seem to fit squarely within the cause-lawyering tradition, under a sense of moral imperative to “do the most good.” Within this prioritist advocacy framework, LPP members regularly deliberated over second-order representational issues. In one meeting, I suggested potential labels for LPP’s theory of accountability, beginning with “first-principle accountability.” This framing stresses moral commitment, although participants noted that the term lacks recognition of their methodology for maximizing impact. One participant suggested “internal accountability.” As opposed to “external accountability,” where “you have direct feedback from living constituencies,” this participant explained that focusing on future generations, who cannot provide input, means that this movement must find accountability through the integrity of their own methodologies and practices.

Regarding the commitment to the interests of future generations, I proposed the term, “astronomical value accountability,” in reference to the arguably great value associated with future persons—with a future potential of an estimated $10^{16}$ “human lives of normal duration,” in addition to the likelihood that technological developments could greatly increase this number if we do not destroy ourselves along the way.\textsuperscript{121} Participants were unpersuaded by this label, in part because it seems to

\textsuperscript{117} An internal LPP report notes that a “long-term cause lawyer” may face conflicts of interest where they are forced to “choose between the cause and the client.” The report further notes a desire to “avoid the type of victimization” that has occurred in some civil rights litigation where plaintiffs’ interests have been overridden or “misrepresented.”

\textsuperscript{118} See Stephen L. Pepper, \textit{The Lawyer’s Amoral Ethical Role: A Defense, A Problem, and Some Possibilities}, AM. BAR FOUND. RES. J. 613 (1986) (defending the “standard conception” of the lawyer role, which demands zealous partisanship on behalf of the client’s interests while holding in abeyance the lawyer’s own moral assessments of client objectives).

\textsuperscript{119} \textit{Id.}

\textsuperscript{120} See Cummings, supra note.

\textsuperscript{121} Nick Bostrom, \textit{Existential Risk Prevention as Global Priority}, 4 GLOB POLICY. 15 (2013); BOSTROM AND CIRKOVIC, supra note (detailing scientific predictions that complex life on
imply “strong longtermism,” the view that our greatest priority today should be impacting the far future.\textsuperscript{122} Most participants were uncertain about their support for this brand of longtermism. A less “strong” version of “future generation accountability” might be a more accurate term, reflecting the general sense of compassion and concern that participants expressed for people who will one day exist and have meaningful lives if we can avoid an existential event.

Within this discussion, some participants were skeptical of any suggestion that they “represent” future generations, because this would imply a fiduciary relationship that seems impossible given the inability to receive communications from the parties being represented.\textsuperscript{123} Participants described an “epistemic challenge” when seeking to know what future generations might want and need. History reveals vast changes in human values over centuries and millennia, which suggests that anything that we might lock in today, no matter how beneficial it seems to us, might be deemed undesirable by future generations. One common response to the epistemic challenge is to assume that future generations would want their “basic needs,” as defined in the lower third of Maslow’s hierarchy, such as the necessities of survival and a baseline of well-being.\textsuperscript{124} Another commonly cited response emphasizes the principle of optionality, which advises avoiding locked-in and irreversible effects. Efforts to mitigate existential risk seem to support both of these responses—ensuring basic needs and preserving optionality by avoiding extinction or permanent dystopia. This discussion raises the difficult questions introduced supra Part I regarding whether, and to what extent and under what conditions, human extinction matters. But participants in this study generally saw at least some significant value in preventing human extinction. And they saw great value in avoiding locked-in dystopias. This assignment of value seems to help participants develop a sense of accountability to protect future generations from an existential catastrophe.

A key point of debate that arises in these discussions is how much accountability is owed to current living persons. A few participants saw little need for this form of accountability. As one participant explained, to follow the prevailing norms and values of “present humans” may be “almost irresponsible” because humans generally

---

\textsuperscript{122} Greaves and Macaskill, supra note (defining strong longtermism as “the view that impact on the far future is the most important feature of our actions today”).

\textsuperscript{123} See MACASKILL, supra note (“Though we cannot give genuine political power to future people, we can at least give consideration to them. Abandoning the tyranny of the present over the future, we can act as trustees.”).

\textsuperscript{124} See Saul McLeod, Maslow’s Hierarchy of Needs, 1 SIMPLY PSYCH. 1 (2007).
are subject to cognitive biases against recognizing existential risks and the moral interests of such a distant population as future generations. Another participant argued that conceiving of accountability primarily to future rather than current persons is more inclusive than prevailing theories of democracy, which can be critiqued for “only taking into consideration the interests and preferences of people currently alive.” But most participants took seriously the notion that working on existential risk requires a significant degree of accountability to current persons. According to leading scholars in the existential risk community, people alive today appear to face an alarming probability of existential or sub-existential catastrophe that could cause a great deal of suffering and loss of life.125 Participants’ regularly discussed how these harms, if realized, would likely be distributed unequally, along the familiar axes of global inequality.126 Moreover, recent policy proposals in this field have recommended new governing structures that bring a broader community of current persons into the efforts to protect future persons, e.g. the proposed creation of legislative citizens’ panels with a mandate to protect future generations.127 Such mechanisms could foster a sense of public accountability by serving as a “legitimate voice for the people.”128

In the discussions I observed on current-person accountability, one of the most commonly debated questions was who, among current people, should have a seat at the table, and who seems to be excluded. This discussion was often framed as a matter of diversity. Although this movement started with an Oxford-based conversation among mostly white men, the existential risk community and the broader effective altruist community are diversifying over time.129 Diversity is defined in a number of ways in this field, e.g. one of the major funders recently called for proposals that address “racial, gender, geographical, ideological, and educational diversity.”130 Relative to many other organizations in this field, LPP appears to be

125 See supra Part I.

126 An internal LPP report emphasizes the inequalities often associated with catastrophic events, as some harms “may be readily avoided and mitigated locally by those with resources.”

127 MacAskill and John, supra note (proposing citizens’ panels as a “novel representative, deliberative, and future-oriented body” with “an explicit mandate to represent the interests of future generations”).

128 Id.

129 See David Moss, EA Survey 2020: Demographics, https://forum.effectivealtruism.org/posts/ThdR8FzcfA8wckTJi/ea-survey-2020-demographics (reporting a survey of the effective altruist community showing that respondents disproportionately identify as men (70.5%), white only (75.9%), and young (median age of 27, mean 29, 80% younger than 35, but noting that these percentages appear to be declining over time).

130 This is a quotation from the Future Fund call for proposals issued in January 2022.
quite diverse in terms of international representation. In addition to their wide-ranging geographic backgrounds, LPP adopts a cultural value to “think globally,” which is reflected in their strategic discussions (e.g. planning legal interventions according to their cross-jurisdictional impacts), as well as informal conversations (e.g., remarking on cultural differences as revealed through anecdotes from day-to-day life or different reactions to global news events). As an ethnographer and interviewer, the global nature of this movement was a near constant theme in my observations, whether traveling in different countries or engaging in remote meetings, which were often held at odd hours to accommodate different time zones. Nearly all conversations that I encountered in this field were held in English although most participants speak English as a second language. This movement’s remarkable ability to operate on a global level is facilitated by the widespread rise of remote work, and the related technological tools that have developed, accelerated by the COVID-19 pandemic. While this makes for a diverse community in some respects, participants still emphasized major representational deficits, e.g. the community continues to overrepresent the Global North and white men. I will return to this issue in my discussion of these findings infra Part VI.

Diversity is valued in this community primarily for the sake of inclusivity, i.e. to open the discussion around existential risk to a wider array of voices and interests. But participants also noted that by expanding their current-person representation, diversity efforts may also help make their movement more effective. A more heterogeneous community may have perspectives and information that would otherwise be overlooked, and this may serve the movement’s goal of maximizing impact. Moreover, participants noted that cultivating a broader range of voices may be essential to persuading lawmakers and policymakers to support important actions relating to existential risk. For example, in the international policy space participants observed that some Global South diplomats are particularly resistant to proposals relating to existential risk, both because such proposals seem to divert attention away from issues currently affecting their constituencies and due to general distrust of the Global North countries where theories of existential risk have originated. One response to this issue may be found in LPP’s support for the development of new existential risk initiatives and effective altruism fellowships in the Global South.\footnote{This includes initiatives in Kenya and Mexico and discussions of initiatives in several other countries.} For example, an LPP member from Kenya recently led an 11-week fellowship with a group of around 60 fellows (mostly law students) in Nairobi. Although this participant reported that some fellows initially found longtermism “almost ridiculous...given the problems we have today of poverty and hunger...and corruption,” his survey at the end of the term showed that the fellows’ “minds were changed drastically” and they rated concerns for “far future people” as the cause area
deserving of the greatest concern. This participant observed that several fellows from this program have continued to engage with the topic of existential risk. This may be a promising sign for continued outreach efforts in the Global South. If these efforts are successful, and if the Global South takes a larger leadership role in advocacy around existential risk, participants noted that this may help to create a more effective and persuasive demand for legal and political action.

V. THE CULTURE OF PRIORITIST ADVOCACY

Having discussed the prioritist theories of efficacy (maximizing moral impact) and accountability (with a focus on representing the interests of future generations), this Part examines how such a model is realized in the daily culture of social-change lawyering. One might assume that prioritist advocacy could serve as an ideal but would find little expression in practice. This model demands a commitment to evidence-based reasoning, while setting aside other considerations and incentives. It also demands a commitment to working on behalf of a future population that is invisible and difficult to even imagine. Yet, at least in this early stage of their movement, participants have established a remarkable culture of adhering to their methodology and their focus on future generations. This culture is marked by an effort to continually reinforce scientific truth-seeking norms and to limit cognitive biases. As categorized in the findings presented below, these norms relate to uncertainty (Section A), dissent (Section B), emotion (Section C), and identity (Section D).

A. Normalizing Uncertainty

Uncertainty was a salient theme throughout the interviews and ethnographic observations in this study. Like other forms of risk analysis, estimating the probability of existential threats requires best guesses based on available evidence and theory. Moreover, participants expressed uncertainty, both normative and empirical, regarding the notion that existential risk should be a global priority. But, rather than inhibiting action, uncertainty is viewed in this community as an unavoidable aspect of their work and as something to embrace as a cultural value. As one participant put it, there is an effort within this community to “normalize uncertainty” within their own organizations and their outward-facing advocacy and educational efforts. As one participant explained, “anyone who comes across as

---

132 See, e.g., Keiran Harris, Effective Altruism in a Nutshell (Oct. 18, 2021), https://80000hours.org/2021/10/effective-altruism-in-a-nutshell/ (emphasizing the importance of uncertainty and humility within effective altruism, and conceding, “maybe the sceptics are right, and we’re just wasting our time.”).
overly certain will be greeted with suspicion.” This “epistemic humility,” as it was often termed, is an express commitment of the effective altruist framework. For example, posts on the Effective Altruism Forum generally begin with an “epistemic status,” as recommended by the Forum guidelines, describing the degree of confidence the author has in their empirical and normative claims.133 In order to pursue strategic action under conditions of uncertainty, the prioritists look to expected value theory, approaches to low-probability/high-impact events in decision theory, and Bayesian reasoning.134 Under this approach, even an unlikely outcome can be a highly prioritized goal, e.g. through the lens of the effective altruist concept of d “hits based” actions, where one is willing to accept a high risk of failure in pursuit of a greater expected value.135 But participants also acknowledged that too much uncertainty and continual reassessment could lead to inefficiencies. As one participant noted: “I think in the end, one has to commit to a certain path at least for some time... [rather than] changing your trajectory every few weeks.”

B. Embracing Dissent

These norms of epistemic humility depend on a healthy culture of dissent, so that priorities and strategies can be challenged and subject to update and revision. The prioritists draw on the effective altruist commitment to limiting “value alignment” and avoiding the formation of a “monoculture,” which would tend to promote motivated reasoning.136 The Effective Altruism Handbook describes a foundational commitment to “criticism, whether from inside or outside of the movement.”137 The development of strong dissenting arguments is particularly prized in this culture, as

133 A typical example reads: “Epistemic Status: Quickly written (~4 hours), uncertain. [This topic] is not my field of expertise.” These notes are typically accompanied by symbols, e.g. (+) (++ (-) (-) (? (??), which depict degrees of uncertainty and other valences.

134 See Bostrom, supra note; MACSKILL, supra note.

135 See Holden Karnofsky, Hits-based Giving (Apr. 4, 2016), https://www.openphilanthropy.org/blog/hits-based-giving (“One of our core values is our tolerance for philanthropic ‘risk.’ Our overarching goal is to do as much good as we can, and as part of that, we’re open to supporting work that has a high risk of failing to accomplish its goals. We’re even open to supporting work that is more than 90% likely to fail, as long as the overall expected value is high enough.


137 Cotra, supra note; see also, MacAskill, supra note (noting that the effective altruist culture of dissent and “independence of thought” is reflected in the observation that several of the “most-upvoted Forum posts” are “critical” or “critically self-reflective” in nature).
evident in reward functions for arguments and evidence that tend to disconfirm effective altruist ideas and conclusions (e.g., “red team challenges” that give prizes to the best criticism and counter-arguments of effective altruist ideas).\textsuperscript{138} During the period of ethnographic observations, LPP conducted surveys, held discussions, and wrote a report on the topic of honest feedback and “normalizing being more critical toward each other’s work.”\textsuperscript{139} A participant at LPP noted that they were hired for a research-focused position “with the goal of advancing longtermism,” but explained that “advancing” was meant in a scientific sense, which includes developing counterarguments and identifying uncertainties, and thus “pursuing the truth, be it favorable to longtermism or not...through science and research.”

Given this commitment to continually expressing misgivings and objections, it is perhaps surprising that this culture is also marked by what LPP labels a “warm, kind, and supportive” environment, which was very much how this culture appeared through the lens of this qualitative study. For example, weekly all-hands meetings began with a round of “achievements and gratitude” where attendees shared personal updates (e.g., openly disclosing details about health, childcare, recreation, pets, and family) and expressed appreciation for various things in their lives including assistance and feedback they received from colleagues in the meeting. LPP members also held recurring sessions (every one to two weeks) focused on mental health, team-building activities, “informal hangouts,” and one-on-one “watercooler” meetings. When someone spoke in an online meeting, they were frequently greeted with supportive emojis (e.g. hands clapping, party hat, hearts, and crying laughing). Crucially, this support was most often directed toward one’s reasoning or articulation of different positions, not necessarily their conclusion, which generally remained subject to uncertainty and debate even at the end of a discussion. Some participants suggested that these warm relationships, when paired with a norm of valuing counterarguments, can encourage members to view disagreement as a means to find the best ideas rather than a personal attack. It is possible that this culture could become “too nice,” as one participant noted, such that members could grow unwilling to express dissent out of fear of breaking a norm of amicability. And some online commentaries have expressed the opposite concern, suggesting that effective altruism, and the longtermist community in particular, is inhospitable to

\textsuperscript{138} See e.g., Criticism and Red Teaming Contest, https://forum.effectivealtruism.org/topics/criticism-and-red-teaming-contest (“a contest on the Effective Altruism Forum for writing that critically engages with theory or work from the effective altruism community”).

\textsuperscript{139} In some public-facing materials, LPP has described their cultural commitment to “discuss ideas openly and honestly, letting the best ideas win. Honest criticism and feedback are both welcome and expected.”
certain lines of counterargument.\textsuperscript{140} But, overall, as observed in this study, this culture of what I reference in my field notes as “supportive dissent” was one of the most striking features of prioritist advocacy.\textsuperscript{141} LPP members and other participants in this study showed a consistent willingness to revisit and debate even the most fundamental concepts that motivate their work, as well as the meta-level question of finding the right mix of agreeableness and dissent.

\textit{C. Limiting System-1 Deliberation}

These values of embracing uncertainty and dissent within deliberative decision-making processes could be undermined if participants were driven by intense emotions and automatic, instinctive, “System-1” cognition.\textsuperscript{142} For example, it is conceivable that working on existential risk on a daily basis would engender overwhelming fear about apocalyptic scenarios, anger toward particular entities that exacerbate risks, or hope for visions of a utopian future. But these framings were uncommon and disfavored among participants. Some even referenced a “missing mood,” wherein their concern for the “long-term future of billions and billions of people” is not met by a commensurate emotional response.\textsuperscript{143} This missing mood may be a product of the cognitive biases that fundamentally limit our ability to comprehend the scale of existential risks (as discussed supra Part I). Even for the participants in this study who work with existential risk on a daily basis, it is difficult, as one participant put it, to “emotionalize uncertainty” regarding the likelihood of existential events and to generate strong feelings for “people who don’t exist yet,” because “you can’t meet them” and ‘you can’t have a graphic documentary about them.” One participant contrasted this missing mood with their work in

\textsuperscript{140} See, e.g., Carla Zoe Cremer and Luke Kemp, \textit{Democratising Risk: In Search of a Methodology to Study Existential Risk} (forthcoming). \textit{See also} Lucius Caviola, \textit{Against Naïve Effective Altruism} (Nov. 20, 2017), https://www.youtube.com/watch?v=2oRgxxafXk&ab_channel=CentreforEffectiveAltruism (describing the risk that a naïve understanding of effective altruism could lead to being overly concerned with “signaling contrarianism” and “appearing cool and wise by holding weird beliefs,” and as a result being overly dismissive of common sense).

\textsuperscript{141} By coincidence, I have learned that this term is very similar to what effective altruists call, “supportive scepticism in practice.” \textit{See} Michelle Hutchinson, \textit{Supportive Scepticism in Practice}, https://forum.effectivealtruism.org/posts/CkikpvdkkLLJHhLXL/supportive-scepticism-in-practice.

\textsuperscript{142} \textit{See generally}, DANIEL KAHNEMAN, \textit{THINKING, FAST AND SLOW} (2011) (describing “system 1” cognition as emotional and instinctive (“fast”) and “system 2” cognition as more rational and deliberative (“slow”).

\textsuperscript{143} \textit{See} Musk’s \textit{Non-Missing Mood} Missing mood: http://lukemuehlhauser.com/musks-non-missing-mood/
animal welfare, where advocates would frequently “cry together” during meetings,\(^{144}\) and in social justice contexts, where this participant remarked that they “miss the...bleeding heart of being more emotionally drawn to the cause.”\(^{145}\)

A few participants described their work in this field in more emotional terms. For example, one participant explained that they were drawn to working on existential risk after undergoing extensive psychotherapy to get more “connected to [their] emotions” and to have more “emotional capacity for compassion with broader groups and issues,” which led to thinking about how they could “care about other people in a meaningful way” and then whether there are “more and less effective ways of doing this.” Two lawyer participants reported drawing a poignant source of motivation from the 2020 entreaty of a suicide note written by a Harvard Law Effective Altruism member, who wrote, as reported by family members in the press: “Please look after each other, the animals, and the global poor for me.”\(^{146}\) One participant described their admiration for this late colleague’s deeply felt and expansive compassion and “beautiful heart.”

This issue was sometimes described by participants as a question of rationality. For example, I observed an internal LPP discussion over whether to continue using the term “rational” in the public description of their organizational culture, which, at the time, read: “Rational: We use evidence and careful analysis to tackle the world’s most pressing problems...” Some members worried that this term may imply a simplistic rational/emotional dichotomy and a judgement toward others for being “irrational,”\(^{147}\) although some conceded that the term accurately describes “an interest in being good with thinking, good with statistics,” and an effort to encourage “Bayesian thinking” and “ways of reducing cognitive bias.” This terminological debate reflects a deeper tension regarding the role of emotions in motivating this

---

\(^{144}\) As this participant noted, “seeing animals suffer is emotionally extremely charged, and seeing animals happy is also emotionally charged.”

\(^{145}\) Another participant similarly noted, “I think it is infrequent that I’m super connected to the emotions of ‘oh no, how horrible would it be if this [catastrophic event] really happened?’ Compared to being involved in the social justice movement where I can see these things and experience them in my whole life...”


\(^{147}\) See James M. Jasper, Emotions and Social Movements: Twenty Years of Theory and Research, 37 ANN. REV. SOCIOLOGY 1 (2011) (challenging the rationality/emotion dichotomy with the observation that “feeling and thinking are parallel” and are “composed of similar neurological building blocks”).
community. Effective altruism is often said to require “not going with your gut” and instead “taking a step back, figuring out your values, and determining where you can have the most impact.”

The effective altruism framework scrutinizes emotional reasoning, in recognition that often “choosing from our heart is unfair,” and it is crucial to “do the math, to go find the numbers...figure out how many people a problem affects...[and] how badly it affects them...” In Doing Good Better, MacAskill recommends “combining the heart and the head,” although he wrote this at a time when current-people causes (e.g., global health and animal welfare) were generally favored by the community’s prioritization methods. Combining the heart and the head might be more difficult in the context of existential risk, which raises the less emotionally salient issue of protecting future generations. Some participants sought to resolve this tension by finding “outlets for the heart” in their pro bono and philanthropic contributions outside of their full-time daily efforts to reduce existential risk. One such participant explained that they feel more “emotional connection” to their donations to effective global poverty interventions, because: “my heart is there [with global poverty], and my head is with longtermism.”

D. Limiting Identification

The majority of participants in this study expressed considerable hesitation when describing themselves as “effective altruists,” “longermists,” or “advocates for existential risk mitigation.” As one participant explained, identifying too strongly with such labels could imply a “movement [with] core tenets you have to follow,” whereas they conceived of themselves belonging to a “scientific community” marked by heterogeneous discourse, dissent, and uncertainty. These scientific norms may make it difficult to conceive of solidarity around a shared identity. Moreover, participants suggested that a strong sense of solidarity would be undesirable because it would create “social pressure to conform,” a preference for in-group members, and, as one participant put it, an “us vs. them solidarity and antipathy.”

148 MacAskill, supra note at 10 (arguing that “relying on good intentions alone to inform your decisions is potentially disastrous”).

149 Cotra, supra note.

150 MacAskill, supra note (noting that the effective altruist notion of “combining the heart and the head” can mean that the heart inspires altruistic pursuits and the head informs those pursuits by turning “good intentions into astonishingly good outcomes”).

151 One participant noted that effective altruism does not particularly lead to identification because it is a “very diverse movement” full of “different views and priorities” and a strong commitment to debate. See also, Ajeya Cotra, supra note (“I say that I’m an effective altruist. That just means a person trying to be effective at altruism.”).

152 See Benjamin Todd (@ben_j_todd), Twitter (Aug. 8, 2021, 3:54PM) https://twitter.com/ben_j_todd/status/1424458937286512647 (“turning effective
concerns have been raised in sociological research finding that members of social movements tend to develop a strong preference for in-groups, while undergoing “identity work” to “preserve or to enhance their egos” by aligning their views with the goals of the movement. Social movements may tend to limit available “argument pools” due to dynamics of “enclave deliberation.” Participants in this study cited psychological research on “moral tribes” and “righteous minds,” widely

altruism into an identity has been powerful, but has had many downsides,” and this includes that it “creates social pressure to conform”); Lizka, Against Longtermist as an Identity, EA FORUM (May 13, 2022) https://forum.effectivealtruism.org/posts/FkFTXKeFxwcGiBTwkg/against-longtermist-as-an-identity (arguing that identifying as a longtermist can “make it harder to change your mind based on new information” and can lead to confusion where the “group identity” encourages viewpoints that one would “otherwise not have adopted as part of your individual identity.”); Helen Effective Altruism is a Question (not an ideology), EA FORUM (Oct. 16, 2014) https://forum.effectivealtruism.org/posts/FpjQMYQmStWewZ83/effective-altruism-is-a-question-not-an-ideology (arguing against identifying as an effective altruist in favor of stating that one is an “aspiring effective altruist,” a “member of the Effective Altruism movement,” or “interested in Effective Altruism”); see also Cullen O’Keefe, Proposed Longtermist Flag, EA FORUM, https://forum.effectivealtruism.org/posts/efd4B2LLd3DXGiVSP/proposed-longtermist-flag (Mar. 24, 2021) (suggesting the possibility of creating a flag for longtermism, as has proven useful in “mature and successful movements,” but raising the concern that a flag might encourage partisan psychology and “ideological loyalty,” and that it could “stymie open and honest discourse about longtermism, including criticism thereof.”).

Cantril, supra note at 162 (observing that, upon joining a social movement, “the individual is now an in-group member of a rather highly selected gathering,” and that such identification with the movement can cause the individual to “lose themselves in some causes”); Jasper, supra note (suggesting that movements tend to “minimize affective loyalties to anyone outside the group and maximize them to the group or its leaders,” and that movements tend to frame in-groups positively relative to out-groups); Snow and McAdam in Self, Identity, and Social Movements, at 230 and 23 (summarizing a body of literature suggesting that social movements are especially appealing to those with a “spoiled” identity and thus engaged in a “collective search for identity”).

Sunstein, supra note (observing that when one conceives of oneself “as part of a group having a degree of connection and solidarity...group polarization is all the more likely, and it is also likely to be more extreme...” and noting that these dynamics “tend to suppress dissent and thus to lead to inferior decisions.”).

read in the effective altruism community, suggesting that group identification tends to inhibit truth-seeking and moral reasoning. A recurring theme in my field notes was the surprising (to me) absence of judgment and derision toward non-prioritists. I had anticipated that a group that works to set optimal priorities might view the rest of the world as setting the “wrong” priorities (e.g. failing to address existential risk) and thus would tend to define themselves in opposition to these other groups. But over the course of this study such judgment was rarely expressed and did not appear to be a strong source of motivation.

But this norm against group identification should not be exaggerated. Participants acknowledged some benefits of aligning personal and collective identities—under a notion similar to the “identity/movement nexus” cited by sociologists as a crucial element in the formation and growth of social movements.¹⁵⁶ Some participants described their affinity with prioritist colleagues in terms of a common “life plan,”¹⁵⁷ a set of background readings, and a general commitment to “thinking about making a fair world” through an evidence-based methodology.¹⁵⁸ This creates a basis for some degree of shared identity, although this identity is rooted more in their commitment to their methodology and epistemic culture rather than any particular conclusions about what cause areas or strategies should be most prioritized. As one participant summarized this point: “We unite around our epistemic norms.” Thus, in my field notes, I referenced this concept as “epistemic identification.”

toward group-based morality, which leads to a failure to appreciate valid perspectives among out-groups).

¹⁵⁶ Snow and Mcadam, supra (citing the example of the mid-twentieth century Civil Rights Movement, where movements leaders drew on church identity, in addition to racial identity, to mobilize and foster commitment to the cause, while also transforming identities of activists in the Freedom Summer). See David Snow, Collective Identity and Expressive Forms (2001) (discussing mechanisms of identity convergence in social movements, including amplification (building on existing identities that are already congruent with the movement), consolidation (revealing the compatibility of identities that seemed inconsistent), extension (making existing identities more pervasive or salient), and transformation (forming new identities)); Cass Sunstein, How Change Happens 2019 (noting that group polarization has been useful to spur movements, citing a number of examples including feminism, the Civil Rights Movement, and Reaganism).

¹⁵⁷ See Ajeya Cotra, supra note (noting that effective altruism presents a “radical but very simple plan for your life. Figure out how to do the most good and then do it.”).

¹⁵⁸ This common set of readings and other effective altruism materials was evident, for example, in the application questions for a recent reading group on law and effective altruism, which asked if applicants had attended effective altruist events, listened to relevant podcasts, read career advice from 80,000 Hours, read the Effective Altruism Handbook, or read writings from a list of effective altruist and longtermist theorists, including Will MacAskill, Nick Bostrom, and Toby Ord.)
This form of identity may not be the most powerful basis for motivation and recruitment. One participant contrasted their relatively “weak” identification as an effective altruist with their past experience in social justice movements (including attending demonstrations for criminal justice reform and several “leftist” causes), which fostered feelings of “certainty and a sense of belonging and having a clear enemy.” Epistemic identification may fail to provide quite the same “warm glow,” as this participant put it, of belonging and collective voice. But this same participant noted that while they were engaged in other social movements, what was missing was a “focus on reasoning,” which they were grateful to find in the prioritist community. Other participants similarly noted that they had been generally disappointed by the lack of evidence-based reasoning and high epistemic standards in their experiences in academia, legal practice, government service, and in society generally. As already noted, these participants did not tend to denigrate these out-group contexts. Instead, they emphasized their sense of appreciation for the effective altruist and longtermist communities, particularly when they offered biographical accounts of first discovering these communities and “finding a home” when they realized the shared commitment to, as one participant put it, “take epistemics seriously.”

In sum, this analysis suggests a tension within the prioritist advocate identity—which offers a sense of belonging but also demands that one resist overly identifying with the group. Several participants noted that the shared experience of this tension can foster a sense of solidarity, whereby members identify with one another on the basis of the shared norm against over-identification. This form of solidarity is grounded in a contradiction (group identity based on resisting group identity), and so is inherently limited but perhaps to a desired degree. Some participants suggested that maintaining this identity tension is crucial to realizing the ideals of prioritist advocacy in practice. Too little identification with movement and cause might lead to a lack of personal motivation. But too much identification might tend to undermine the cultural commitments to uncertainty, dissent, and deliberative decision-making.

VI. DISCUSSION

In his 2004 book, *Catastrophe: Risk and Response*, Professor Richard Posner argued that the legal profession had failed to play its crucial role in mitigating catastrophic risks, and that this was largely due to the profession’s norms against the scientific method and mathematical reasoning.159 This criticism took particular aim at lawyers for their “culture of advocacy and doctrinal manipulation,” wherein they assume the

159 POSNER, supra note (observing that law lacks “empirical methodology” and “error correction,” and is “more like a language than a science”).
truth (e.g. interpretations of the law or facts that are favorable to their client) and then seek to persuade others of this truth rather than engaging in any sort of scientific process for evaluating truth claims.\textsuperscript{160} Posner contrasts the scientific orientation “toward knowledge,” where “propositions are accepted only if they survive confrontation with experimental data,” with the law’s orientation “toward action,” “bending the rules...fitting them to goals,” and asserting certitude, by, for example, declaring “my client is innocent, and that’s the truth.”\textsuperscript{161} Posner’s advises law schools to recruit more students with STEM backgrounds and provide basic STEM education within the law curriculum. This would then help produce a class of “catastrophic risk lawyers” who appreciate the prerequisite scientific methods, probabilistic claims, and decision theory. Posner’s analysis would seem to suggest a culture clash between the notion of learning to “think like a lawyer”\textsuperscript{162} and what the prioritists call learning to “think like an effective altruist.”\textsuperscript{163} If such a clash exists, the participants in this study are living evidence of the exceptional case that Posner hoped to foster—a group of lawyers and other legal advocates who are addressing large-scale risks through an express commitment to scientific reasoning and truth-seeking.

Why have the participants in this study developed such a distinctly scientific understanding of their legal activism? Most participants do not have the STEM backgrounds emphasized by Posner. Instead, they draw from the tradition of effective altruism, which demands evidence-based reasoning. But it is also possible that the prioritist approach observed in this study is greatly influenced by the nature of existential risk as a cause area. Existential risk is a unique kind of issue as the

\textsuperscript{160} POSNER, supra note. This concern is reflected in the laments of legal empiricists that the profession disregards empirical research. See James D. Greiner, The New Legal Empiricism & Its Application to Access-to-Justice Inquiries, 148.1 DAEDALUS 64 (2019) (arguing that the legal profession is “not evidence-based in the scientific sense” and instead tends to “rely on gut intuition and instinct, not on rigorous evidence,” and advocating for the “new legal empiricism,” which has the potential to “transform the U.S. legal profession into an evidence-based field”). This concern also finds support among critics of the role-differentiated morality of lawyers’ professional identities, as the professional role demands zealous partisan advocacy for clients with relatively little opportunity for the lawyer to insert their own political, moral, emotional, and other contextual considerations. This “hired gun” loyalty to clients is limited by professional duties to the public and a blanket requirement to not engage in dishonesty or perjury, but these duties are rarely interpreted to suggest anything like a commitment to scientific methods or reasoning. See generally Richard Wasserstrom, Lawyers as Professionals: Some Moral Issues, 5.1 HUMAN RIGHTS 1, 23 (1975);

\textsuperscript{161} POSNER, supra note ( “the idea of subjecting a legal proposition to a decisive experiment...horrifies the lawyer.”).

\textsuperscript{162} Mertz on legal epistemology (amoral, apolitical, acontextual, unemotional).

\textsuperscript{163} MACASKILL, supra note.
subject of a legal movement. The examples of social-change lawyering that have been studied in the empirical literature generally involve grassroots social movements that are addressing issues of public controversy, often seeking to change cultural norms that are pervasive in daily life (e.g., how we relate to one another across lines of race, gender, and sexuality). In contrast, the activists examined in this article are addressing an issue that is not framed as a matter of widespread norm-shifting and collective organizing. It would be difficult to even conceive of mobilizing large populations to form a vociferous social movement around low-probability/high-impact harms, particularly where such harms are thought to primarily endanger a distant population of future generations. This lack of what socio-legal scholars call “mobilizing frames” may help to explain why the advocates working on this issue find value in the prioritist model.\footnote{See David A. Snow, Rens Vliegenthart, and Pauline Ketelaars, \textit{The Framing Perspective on Social Movements: Its Conceptual Roots and Architecture}, 77 THE WILEY BLACKWELL COMPANION TO SOCIAL MOVEMENTS 392 (2019).} A relatively small expert-based community of activists, which has not attracted much public controversy, may be especially well-suited to a deliberative, scientific methodology for designing legal and other strategies.

One might speculate that prioritist advocacy is, to some extent, a reflection of privileged identities (white, male, elite), people who have less appreciation for current suffering and oppression in the world and thus tend to place a greater focus on a more distant population of future people. Moreover one might associate these privileged identities with some of the cultural traits of the prioritist method—less group identification, less emotional connection to their work, and less focus on local action. These speculations should be subjected to empirical inquiry, but they may suggest support for this movement’s efforts to diversify their community. As discussed supra part IV, diversification is valued by advocates in this field because it may lead to greater inclusivity and legitimacy, as well as more accurate prioritist calculations. Recent scholarship on public-interest law has noted that the prevalence of white leadership in movements for civil rights has led these movements to overlook key injustices and other considerations.\footnote{See Atinuke O. Adediran & Shaun Ossei-Owusu, \textit{The Racial Reckoning of Public Interest Law}, 21 CAL. L. REV. (2021) (calling for greater scrutiny of the racial composition of U.S. public interest law as it impacts marginalized communities).} Similarly, the existential risk community would likely benefit from the perspectives of people who experience great injustice and who live in a dystopic reality today—analogous to the permanent dystopian scenarios contemplated by existential risk scholars.

This commitment to diversification may imply a need to expand the existential risk community and reach out to a broader public. In some respects, this community is beginning to pursue a higher-profile approach. This has included public-facing
books, videos, blogs, and podcasts, as well in-depth New Yorker profiles of the leading figures. In August 2022, the release of Will MacAskill’s book, *What We Owe the Future*, received favorable coverage in many major new outlets, including a Time Magazine cover story. The “integrated advocacy” recommendations of recent social-legal literature (see supra Part I.B) would tend to support this turn to broader audiences. These scholars highlight how legal activists in other social movements appear to have found greater efficacy and accountability by “going broad,” embedding their legal work within larger movements and seeking to shape public opinion. This approach is thought to help promote lasting social change by creating a collective demand for reform, which motivates legal and political action and diminishes backlash to movement gains. Just how far should the existential risk community take this received wisdom from the literature? Should they recruit new members with the widest possible net, expanding their movement to create a more powerful collective voice? Should they focus on developing favorable public opinion—e.g. so that people generally understand the notion of existential risk and would support interventions as opposed to viewing the issue as a matter of science fiction or billionaires’ whims? Or should this community stay relatively narrow with a focus on putting experts in conversation with powerful decisionmakers, including politicians and judges (e.g. perhaps scaling back the plans for record-setting


169 See, CUMMINGS, supra note.
campaign contributions, engagement with heads of state at the UN Summit of the Future, and legal interventions that touch on matters of public controversy)?

These questions are beyond the scope of this article, although one consideration bears directly on this study’s core findings: Efforts to expand this movement may tend to compromise the prioritist culture and methodology outlined in this article. For example, in order to recruit more broadly, this community may tend to replace its “epistemic identification” (i.e., “we unite around our epistemic norms”) with a more conventional approach to strongly identifying with movement, cause, and strategic solutions. In other words, they may tend to adopt a collective identity based on shared conclusions rather than shared questions and truth-seeking methods. Broader movement building could also require stirring up fear and other strong emotions around existential risk. Some organizations in this field have explored this approach by producing short films about “slaughterbot” scenarios, where swarms of small autonomous drones execute people with great efficiency. Participants in this study strongly disfavored these videos on the grounds that, as discussed above (supra Part V.3), motivations rooted in powerful emotions may cause the movement to drift toward a focus on near-term and smaller-scale catastrophic risks, while losing sight of existential threats to the entire future of humanity.

At least at this early stage, prioritist advocacy seems very effective at avoiding this drift. The prioritists are remarkably consistent in their methodology and their cultural commitments, assessing nearly all strategic decisions according to how much each option can be expected to reduce our overall level of existential risk. This ability to “keep your eyes on the prize,” as activists in the Civil Rights Movement used to sing,170 is a defining feature of the existential risk movement as currently constituted. As this community scales up and pursues more direct legal interventions, their culture will likely need to adapt and face some compromise. But maintaining some degree of relatively uncompromised prioritist methodology at the core of this movement could be vital to the movement’s ultimate success. Political incentives threaten to steer the movement toward issues of popular concern among voters, markets, and movements. And, as humans, we face cognitive biases that limit our ability to appreciate existential risk, an issue that is so uncertain, unprecedented, large-scale, and seemingly remote—affecting future generations who are an abstraction well beyond our usual moral circles. Overcoming these incentives and biases is no small task. Prioritist advocacy proposes a framework of continual evidence-based re-evaluation backed by strong cultural commitments to scientific reasoning. In this way, the prioritists seek to keep their focus on representing the future generations whose well-being and existence hang in the balance.

---

170 This is a lyric from the traditional African American spiritual, “Gospel Plow.” See DUKE ELLINGTON & MAHILIA JACKSON, Keep Your Hand on the Plow, on LIVE AT NEWPORT 1958 (Columbia Records 1958).