**AT: Framing**

**AT: Probability First**

**Prefer util and consequences---castigating calculation as cold-blooded sacrifices the infinite value of every human life at the altar of sanctimonious personal purity---shut up and multiply**

**Yudkowsky 15** [Eliezer Yudkowsky, autodidact, co-founder and research fellow at the Machine Intelligence Research Institute (MIRI), founder of LessWrong, and author of the all-time most popular Harry Potter fanfiction, “Feeling Moral,” March 11, 2015, <https://www.lesswrong.com/s/waF2Pomid7YHjfEDt/p/Nx2WxEuPSvNBGuYpo>]

You know what? **This isn’t about your feelings. A human life, with all its joys and all its pains**, **adding up over the course of decades, is worth far more than your brain’s feelings of comfort or discomfort with a plan**. **Does computing the expected utility feel too cold-blooded for your taste?** **Well, that feeling isn’t even a feather in the scales, when a life is at stake.** Just shut up and multiply. A googol is 10^100—a 1 followed by one hundred zeroes. A googolplex is an even more incomprehensibly large number—it’s 10^googol, a 1 followed by a googol zeroes. Now pick some trivial inconvenience, like a hiccup, and some decidedly untrivial misfortune, like getting slowly torn limb from limb by sadistic mutant sharks. If we’re forced into a choice between either preventing a googolplex people’s hiccups, or preventing a single person’s shark attack, which choice should we make? If you assign any negative value to hiccups, then, on pain of decision-theoretic incoherence, there must be some number of hiccups that would add up to rival the negative value of a shark attack. For any particular finite evil, there must be some number of hiccups that would be even worse.

Moral dilemmas like these **aren’t conceptual blood sports** for keeping analytic philosophers entertained at dinner parties. **They’re distilled versions of the kinds of situations we actually find ourselves in every day.** Should I spend $50 on a console game, or give it all to charity? Should I organize a $700,000 fundraiser to pay for a single bone marrow transplant, or should I use that same money on mosquito nets and prevent the malaria deaths of some 200 children?

Yet **there are many who avert their gaze from the real world’s abundance of unpleasant moral tradeoffs**—many, too, who take pride in looking away. Research shows that people distinguish “sacred values,” like human lives, from “unsacred values,” like money. When you try to trade off a sacred value against an unsacred value, subjects express great indignation. (Sometimes they want to punish the person who made the suggestion.)

My favorite anecdote along these lines comes from a team of researchers who evaluated the effectiveness of a certain project, calculating the cost per life saved, and recommended to the government that the project be implemented because it was cost-effective. The governmental agency rejected the report because, they said, you couldn’t put a dollar value on human life. After rejecting the report, the agency decided not to implement the measure.

Trading off a sacred value against an unsacred value feels really awful. To merely multiply utilities would be too cold-blooded—it would be following rationality off a cliff . . . But **altruism isn’t the warm fuzzy feeling you get from being altruistic.** If you’re doing it for the spiritual benefit, **that is nothing but selfishness. The primary thing is to help others, whatever the means. So shut up and multiply**!

**“Infinite risks,” “anti-knowledge,” and paralysis are wrong** --- even if assigning risk is hard, the alternative is to forego it entirely

**Rendall 17** [Matthew Rendall, Lecturer in Politics and International Relations at the University of Nottingham Political Science degree from Columbia University, “PASCALIAN WAGERING AND CATASTROPHIC RISK: THE WEITZMAN-NORDHAUS DEBATE”, 5/9/17, <https://www.nottingham.ac.uk/climateethicseconomics/documents/papers-workshop-5/rendall.pdf>]

Jonas, however, feared not only the physical destruction of humanity but also losing ‘the essence of man’. His imperative also forbids taking any risks with the latter.37 Depending on how we interpret ‘the essence of man’, this might give it a much broader scope.38 Indeed, Jonas maintained that there were many “apocalyptic” threats, whether nuclear weapons or nuclear reactors.39 Interpreted that way, the imperative may **seem to threaten** the **same paralysis** as does **Pascal’s Wager**.40 Nordhaus warns that if we attached infinite disutility to existential risks, it could be rational to fight preventive wars to prevent barely discerned power shifts in the far future.41 If the imperative of responsibility is to avoid paralysis and paranoia, its scope must be restricted to a small number of truly apocalyptic risks.42

While it is true that nearly anything we do could **in principle** prove catastrophic, in most cases we have no more reason to believe that acting is riskier than abstaining. It is often more dangerous, for example, to launch preventive wars than to forego them. With climate change, on the other hand, we have a well-theorized causal mechanism, with good theoretical and empirical reasons for believing that higher greenhouse gas concentrations are more dangerous than lower ones.43 There are a small number of other apocalyptic threats that fall into this category—notably thermonuclear war—but for now there are not scores or hundreds.44

Still, there might be many actions for which we can construct a theoretically possible, albeit improbable, story of how they could lead to doom, and that seem slightly more likely than not to bring on the apocalypse. To adapt an example of Jon Elster’s, it is conceivable that commercial TV transmissions will attract visitors from outer space. While such extraterrestrials might either save the earth from some other threat or destroy it, it is unlikely that these probabilities exactly cancel each other out. In the absence of **reliable information**, it is **rational** to make use of **whatever information** we do have—**even if** it is merely a flimsy analogy or a vague hunch. Suppose we judge—very tentatively—that the chance that alien visitors will destroy the earth is slightly greater than that they will destroy it. Should this be a decisive consideration in whether or not to broadcast The Simpsons?45

Intuitively, the question seems absurd, but we should not reject it out of hand. In other contexts when policies appear to carry **a small chance** of **apocalyptic destruction** they should be taken **seriously**, and sometimes are. It was reasonable for Compton to take into account the chance that the atomic bomb would ignite the earth’s atmosphere **even if** it was only on the order of **three in a million**. It is **reasonable** today to consider the possibility that particle accelerators might destroy the earth in deciding whether to run them.46 The TV transmission question may seem absurd because we are by now accustomed to television broadcasts, and they have become an integral part of modern societies. That does not show that they are safe. It may instead indicate that **the industrial revolution** has led to **the proliferation** of **existential risks**. The Stern Review estimated **a ten percent chance** of **human extinction per century**. Surely no British government before the twentieth century, had it considered the matter at all, would have assessed the risk as nearly this high.47

Even if the number of apocalyptic threats to which we must respond is presently circumscribed, with technological development it is likely to grow. In that case the imperative of responsibility could become far more confining.48 Yet in a nightmare world where existential risks had become common, what could we reasonably do but exercise extreme caution? Perhaps the problem is not that the imperative of responsibility is too restrictive, but that technological development is too dangerous. We should not take for granted that industrialization story has a happy ending.49 If **century after century** we continue to run **even small existential risks**, sooner or later, we are bound to lose the gamble.50

**Trust yourself- selectively accounting for certain biases results in worse forms of bias**

**Yudkowsky, ‘8** Eliezer Yudkowsky, co-founder and research fellow at the Machine Intelligence Research Institute 2008, “Cognitive Biases Potentially Affecting Judgment of Global Risks”, In Global Catastrophic Risks, edited by Nick Bostrom and Milan M. Ćirković, New York: Oxford University Press, <https://intelligence.org/files/CognitiveBiases.pdf>, EO

Ironically, Taber and Lodge’s experiments confirmed all six of the authors’ prior hypotheses. Perhaps you will say: “The experiment only reflects the beliefs the authors started out with—it is just a case of confirmation bias.” If so, then by making you a more sophisticated arguer—by teaching you another bias of which to accuse people—I have actually harmed you; I have made you slower to react to evidence. I have given you another opportunity to fail each time you face the challenge of changing your mind.

Heuristics and biases are widespread in human reasoning. Familiarity with heuristics and biases can enable us to detect a wide variety of logical flaws that might otherwise evade our inspection. But, as with any ability to detect flaws in reasoning, this inspection must be applied evenhandedly: both to our own ideas and the ideas of others; to ideas which discomfort us and to ideas which comfort us. Awareness of human fallibility is dangerous knowledge if you only remind yourself of the fallibility of those who disagree with you. If I am selective about which arguments I inspect for errors, or even how hard I inspect for errors, then every new rule of rationality I learn, every new logical flaw I know how to detect, makes me that much stupider. Intelligence, to be useful, must be used for something other than defeating itself.

**They have the burden of rejoinder- if our arguments are so bad they should be able to easily defeat them**

**Yudkowsky, ‘8** Eliezer Yudkowsky, co-founder and research fellow at the Machine Intelligence Research Institute 2008, “Cognitive Biases Potentially Affecting Judgment of Global Risks”, In Global Catastrophic Risks, edited by Nick Bostrom and Milan M. Ćirković, New York: Oxford University Press, <https://intelligence.org/files/CognitiveBiases.pdf>, EO

Every true idea which discomforts you will seem to match the pattern of at least one psychological error.

Robert Pirsig said: “The world’s biggest fool can say the sun is shining, but that doesn’t make it dark out.” If you believe someone is guilty of a psychological error, then demonstrate your competence by first demolishing their consequential factual errors. If there are no factual errors, then what matters the psychology? The temptation of psychology is that, knowing a little psychology, we can meddle in arguments where we have no technical expertise—instead sagely analyzing the psychology of the disputants.

If someone wrote a novel about an asteroid strike destroying modern civilization, then someone might criticize that novel as extreme, dystopian, apocalyptic; symptomatic of the author’s naive inability to deal with a complex technological society. We should recognize this as a literary criticism, not a scientific one; it is about good or bad novels, not good or bad hypotheses. To quantify the annual probability of an asteroid strike in real life, one must study astronomy and the historical record: no amount of literary criticism can put a number on it. Garreau (2005) seems to hold that a scenario of a mind slowly increasing in capability, is more mature and sophisticated than a scenario of extremely rapid intelligence increase. But that’s a technical question, not a matter of taste; no amount of psychologizing can tell you the exact slope of that curve.

It’s harder to abuse heuristics and biases than psychoanalysis. Accusing someone of conjunction fallacy leads naturally into listing the specific details that you think are burdensome and drive down the joint probability. Even so, do not lose track of the real-world facts of primary interest; do not let the argument become about psychology.

**Weigh magnitude times probability—“probability first” framing is rooted in psychological biases and leads to mass death**

**Clarke 8** [Lee, member of a National Academy of Science committee that considered decision-making models, Anschutz Distinguished Scholar at Princeton University, Fellow of AAAS, Professor Sociology (Rutgers), Ph.D. (SUNY), “Possibilistic Thinking: A New Conceptual Tool for Thinking about Extreme Events,” Fall, Social Research 75.3, JSTOR]

In scholarly work, the subfield of disasters is often seen as narrow. One reason for this is that a lot of scholarship on disasters is practically oriented, for obvious reasons, and the social sciences have a deep-seated suspicion of practical work. This is especially true in sociology. Tierney (2007b) has treated this topic at length, so there is no reason to repeat the point here. There is another, somewhat unappreciated reason that work on disaster is seen as **narrow**, a reason that holds some irony for the main thrust of my argument here: disasters are unusual and the social sciences are generally **biased** toward phenomena that are frequent. Methods textbooks caution against using case stud- ies as representative of anything, and articles in mainstreams journals that are not based on probability samples must issue similar obligatory caveats. The premise, **itself narrow**, is that the only way to be certain that we know something about the social world, and the only way to control for subjective influences in data acquisition, is to follow the tenets of probabilistic sampling. This view is a correlate of the central way of defining rational action and rational policy in academic work of all varieties and also in much practical work, which is to say in terms of probabilities. The irony is that probabilistic thinking **has its own biases**, which, if unacknowledged and uncorrected for, **lead to a conceptual neglect** of extreme events. This leaves us, as scholars, paying attention to disasters only when they happen and doing that makes the accumulation of good ideas about disaster vulnerable to issue-attention cycles (Birkland, 2007). These **conceptual blinders** lead to a neglect of disasters as "strategic research sites" (Merton, 1987), which results in **learning less** about disaster than we could and in missing opportunities to use disaster to learn about society (cf. Sorokin, 1942). **We need new conceptual tools** because of an upward trend in frequency and severity of disaster since 1970 (Perrow, 2007), and because of a growing intellectual attention to the idea of worst cases (Clarke, 2006b; Clarke, in press). For instance, the chief scientist in charge of studying earthquakes for the US Geological Service, Lucile Jones, has worked on the combination of events that could happen in California that would constitute a "give up scenario": a very long-shaking earthquake in southern California just when the Santa Anna winds are making everything dry and likely to burn. In such conditions, meaningful response to the fires would be impossible and recovery would take an extraordinarily long time. There are other similar pockets of scholarly interest in extreme events, some spurred by September 11 and many catalyzed by Katrina. The consequences of disasters are also becoming more severe, both in terms of lives lost and property damaged. People and their places are becoming more vulnerable. The most important reason that vulnerabilities are increasing is population concentration (Clarke, 2006b). This is a general phenomenon and includes, for example, flying in jumbo jets, working in tall buildings, and attending events in large capacity sports arenas. Considering disasters whose origin is a natural hazard, the specific cause of increased vulnerability is that people are moving to where hazards originate, and most especially to where the water is. In some places, this makes them vulnerable to hurricanes that can create devastating storm surges; in others it makes them vulnerable to earthquakes that can create tsunamis. In any case, the general problem is that people concentrate themselves in dangerous places, so when the hazard comes disasters are intensified. More than one-half of Florida's population lives within 20 miles of the sea. Additionally, Florida's population grows every year, along with increasing development along the coasts. The risk of exposure to a devastating hurricane is obviously high in Florida. No one should be surprised if during the next hurricane season Florida becomes the scene of great tragedy. The demographic pressures and attendant development are wide- spread. People are concentrating along the coasts of the United States, and, like Florida, this puts people at risk of water-related hazards. Or consider the Pacific Rim, the coastline down the west coasts of North and South America, south to Oceania, and then up the eastern coast- line of Asia. There the hazards are particularly threatening. Maps of population concentration around the Pacific Rim should be seen as target maps, because along those shorelines are some of the most active tectonic plates in the world. The 2004 Indonesian earthquake and tsunami, which killed at least 250,000 people, demonstrated the kind of damage that issues from the movement of tectonic plates. (Few in the United States recognize that there is a subduction zone just off the coast of Oregon and Washington that is quite similar to the one in Indonesia.) Additionally, volcanoes reside atop the meeting of tectonic plates; the typhoons that originate in the Pacific Ocean generate furiously fatal winds. Perrow (2007) has generalized the point about concentration, arguing not only that we increase vulnerabilities by increasing the breadth and depth of exposure to hazards but also by concentrating industrial facilities with catastrophic potential. Some of Perrow's most important examples concern chemical production facilities. These are facilities that bring together in a single place multiple stages of production used in the production of toxic substances. Key to Perrow's argument is that there is no technically necessary reason for such concentration, although there may be good economic reasons for it. The general point is that we can expect more disasters, whether their origins are "natural" or "technological." We can also expect more death and destruction from them. I predict **we will continue to be poorly prepared to deal with disaster**. People around the world were appalled with the incompetence of America's leaders and orga- nizations in the wake of Hurricanes Katrina and Rita. Day after day we watched people suffering unnecessarily. Leaders were slow to grasp the importance of the event. With a few notable exceptions, organi- zations lumbered to a late rescue. Setting aside our moral reaction to the official neglect, perhaps we ought to ask why **we should have expected a competent response** at all? Are US leaders and organiza- tions particularly attuned to the suffering of people in disasters? Is the political economy of the United States organized so that people, espe- cially poor people, are attended to quickly and effectively in noncri- sis situations? The answers to these questions are obvious. If social systems are not arranged to ensure people's well-being in normal times, there is no good reason to expect them to be so inclined in disastrous times. Still, if we are **ever** going to be reasonably well prepared to avoid or respond to the next Katrina-like event, **we need to identify the barriers** to effective thinking about, and effective response to, disas- ters. **One of those barriers** is that we do not have a set of concepts that would help us think rigorously about out-sized events. The **chief toolkit** of concepts that we have for thinking about important social events comes from probability theory. There are good reasons for this, as probability theory has obviously served social research well. Still, the toolkit is **incomplete** when it comes to extreme events, especially when it is used as a base whence to make normative judgments about what people, organizations, and governments should and should not do. **As a complement to probabilistic thinking I propose that we need possibilistic thinking**. In this paper I explicate the notion of possibilistic thinking. I first discuss the equation of probabilism with rationality in scholarly thought, followed by a section that shows the ubiquity of possibilis- tic thinking in everyday life. Demonstrating the latter will provide an opportunity to explore the limits of the probabilistic approach: that possibilistic thinking is widespread suggests it could be used more rigorously in social research. I will then address the most vexing prob- lem with advancing and employing possibilistic thinking: the prob- lem of infinite imagination. I argue that possibilism can be used with discipline, and that we **can be smarter** about responding to disasters by doing so.

**Extinction first**

**GPP 17** (Global Priorities Project, Future of Humanity Institute at the University of Oxford, Ministry for Foreign Affairs of Finland, “Existential Risk: Diplomacy and Governance,” Global Priorities Project, 2017, <https://www.fhi.ox.ac.uk/wp-content/uploads/Existential-Risks-2017-01-23.pdf>, Accessed 7/22/2017, Kent Denver-jKIM)

1.2. THE ETHICS OF EXISTENTIAL RISK In his book Reasons and Persons, Oxford philosopher Derek Parfit advanced an influential argument about the importance of avoiding extinction: I believe that if we destroy mankind, as we now can, this outcome will be much worse than most people think. Compare three outcomes: (1) Peace. (2) A nuclear war that kills 99% of the world’s existing population. (3) A nuclear war that kills 100%. (2) would be worse than (1), and (3) would be worse than (2). Which is the greater of these two differences? Most people believe that the greater difference is between (1) and (2). I believe that the difference between (2) and (3) is very much greater. ... The Earth will remain habitable for **at least another billion years**. Civilization began only a few thousand years ago. If we do not destroy mankind, these few thousand years may be only **a tiny fraction** of the whole of civilized human history. The difference between (2) and (3) may thus be the difference between this tiny fraction and all of the rest of this history. If we compare this possible history to a day, what has occurred so far is only a **fraction of a second**.65 In this argument, it seems that Parfit is assuming that the survivors of a nuclear war that kills 99% of the population would eventually be able to recover civilisation without long-term effect. As we have seen, this may not be a safe assumption – but for the purposes of this thought experiment, the point stands. What makes existential catastrophes especially bad is that they would “**destroy the future**,” as another Oxford philosopher, Nick Bostrom, puts it.66 This future could potentially be extremely long and full of flourishing, and would therefore have extremely large value. In standard risk analysis, when working out how to respond to risk, we work out the expected value of risk reduction, by weighing the probability that an action will prevent an adverse event against the severity of the event. Because the value of preventing existential catastrophe is so vast, even a tiny probability of prevention has huge expected value.67 Of course, there is persisting reasonable disagreement about ethics and there are a number of ways one might resist this conclusion.68 Therefore, it would be unjustified to be overconfident in Parfit and Bostrom’s argument. In some areas, government policy does give significant weight to future generations. For example, in assessing the risks of nuclear waste storage, governments have considered timeframes of thousands, hundreds of thousands, and even a million years.69 Justifications for this policy usually appeal to principles of intergenerational equity according to which future generations ought to get as much protection as current generations.70 Similarly, widely accepted norms of sustainable development require development that meets the needs of the current generation without compromising the ability of future generations to meet their own needs.71 However, when it comes to existential risk, it would seem that we fail to live up to principles of intergenerational equity. Existential catastrophe would not only give future generations less than the current generations; **it would give them nothing**. Indeed, reducing existential risk plausibly has a quite low cost for us in comparison with the huge expected value it has for future generations. In spite of this, relatively little is done to reduce existential risk. Unless we give up on norms of intergenerational equity, they give us a strong case for significantly increasing our efforts to reduce existential risks. 1.3. WHY EXISTENTIAL RISKS MAY BE SYSTEMATICALLY UNDERINVESTED IN, AND THE ROLE OF THE INTERNATIONAL COMMUNITY In spite of the importance of existential risk reduction, it probably receives less attention than is warranted. As a result, concerted international cooperation is required if we are to receive adequate protection from existential risks. 1.3.1. Why existential risks are likely to be underinvested in There are several reasons why existential risk reduction is likely to be underinvested in. Firstly, it is a global public good. Economic theory predicts that such goods tend to be underprovided. The benefits of existential risk reduction are widely and indivisibly dispersed around the globe from the countries responsible for taking action. Consequently, a country which reduces existential risk gains only a small portion of the benefits but bears the full brunt of the costs. Countries thus have strong incentives to free ride, receiving the benefits of risk reduction without contributing. As a result, too few do what is in the common interest. Secondly, as already suggested above, existential risk reduction is an intergenerational public good: most of the benefits are enjoyed by future generations who have no say in the political process. For these goods, the problem is temporal free riding: the current generation enjoys the benefits of inaction while future generations bear the costs. Thirdly, many existential risks, such as machine superintelligence, engineered pandemics, and solar geoengineering, pose an unprecedented and uncertain future threat. Consequently, it is hard to develop a satisfactory governance regime for them: there are few existing governance instruments which can be applied to these risks, and it is unclear what shape new instruments should take. In this way, our position with regard to these emerging risks is comparable to the one we faced when nuclear weapons first became available. Cognitive biases also lead people to underestimate existential risks. Since there have not been any catastrophes of this magnitude, these risks are not salient to politicians and the public.72 This is an example of the misapplication of the availability heuristic, a mental shortcut which assumes that something is important only if it can be readily recalled. Another cognitive bias affecting perceptions of existential risk is scope neglect. In a seminal 1992 study, three groups were asked how much they would be willing to pay to save 2,000, 20,000 or 200,000 birds from drowning in uncovered oil ponds. The groups answered $80, $78, and $88, respectively.73 In this case, the size of the benefits had little effect on the scale of the preferred response. People become numbed to the effect of saving lives when the numbers get too large. 74 Scope neglect is a particularly acute problem for existential risk because the numbers at stake are so large. Due to scope neglect, decision-makers are prone to treat existential risks in a similar way to problems which are less severe by many orders of magnitude. A wide range of other cognitive biases are likely to affect the evaluation of existential risks.75

**Prioritize existential risk prevention---it encompasses AND outweighs other threats.**

Dennis **Pamlin &** Stuart **Armstrong 15**, Dennis Pamlin, Executive Project Manager Global Risks, Global Challenges Foundation, and Stuart Armstrong, James Martin Research Fellow, Future of Humanity Institute, Oxford Martin School, University of Oxford, February 2015, “Global Challenges: 12 Risks that threaten human civilization: The case for a new risk category,” Global Challenges Foundation, p.30-93, https://api.globalchallenges.org/static/wp-content/uploads/12-Risks-with-infinite-impact.pdf

2. Risks with infinite impact: A new category of risks “Most risk management is really just advanced contingency planning and disciplining yourself to realise that, given enough time, very **low probability events not only can happen, but they absolutely will happen**.” Lloyd Blankfein, Goldman Sachs CEO, July 2013 1 **Risk = Probability × Impact** Impacts where civilisation collapses to a state of **great suffering** and **do not recover**, or a situation where all human life end, are defined as **infinite** as the result is **irreversible** and **lasts forever**. A new group of global risks This is a report about a limited number of global risks – that can be identified through a scientific and transparent process – with impacts of a magnitude that pose a threat to human civilisation, or even possibly to all human life. With such a focus it may surprise some readers to find that the report’s essential aim is to inspire action and dialogue as well as an increased use of the methodologies used for risk assessment. The real focus is not on the almost unimaginable impacts of the risks the report outlines. Its fundamental purpose is to encourage global collaboration and to use this new category of risk as a driver for innovation. The idea that we face a number of global challenges threatening the very basis of our civilisation at the beginning of the 21st century is well accepted in the scientific community, and is studied at a number of leading universities.2 But there is still no coordinated approach to address this group of challenges and turn them into opportunities for a new generation of global cooperation and the creation of a global governance system capable of addressing the greatest challenges of our time. This report has, to the best of our knowledge, created the first science-based list of global risks with a potentially infinite impact and has made the first attempt to provide an initial overview of the uncertainties related to these risks as well as rough quantifications for the probabilities of these impacts. What is risk? Risk is the potential of losing something of value, weighed against the potential to gain something of value. Every day we make different kinds of risk assessments, in more or less rational ways, when we weigh different options against each other. The basic idea of risk is that an uncertainty exists regarding the outcome and that we must find a way to take the best possible decision based on our understanding of this uncertainty.3 To calculate risk the probability of an outcome is often multiplied by the impact. The impact is in most cases measured in economic terms, but it can also be measured in anything we want to avoid, such as suffering. At the heart of a risk assessment is a probability distribution, often described by a probability density function4; see figure X for a graphic illustration. The slightly tilted bell curve is a common probability distribution, but the shape differs and in reality is seldom as smooth as the example. The total area under the curve always represents 100 percent, i.e. all the possible outcomes fit under the curve. In this case (A) represents the most probable impact. With a much lower probability it will be a close to zero impact, illustrated by (B). In the same way as in case B there is also a low probability that the situation will be very significant, illustrated by (C). Figure 1: Probability density function [FIGURE 1 OMITTED] The impacts (A), (B) and (C) all belong to the same category, ~~normal~~ [common] impacts: the impacts may be more or less serious, but they can be dealt with within the current system. The impacts in this report are however of a special kind. These are impacts where everything will be lost and the situation will not be reversible, i.e challenges with potentially infinite impact. In insurance and finance this kind of risk is called “risk of ruin”, an impact where all capital is lost.5 This impact is however only infinite for the company that is losing the money. From society’s perspective, that is not a special category of risk. In this report the focus is on the “risk of ruin” on a global scale and on a human level, in the worst case this is when we risk the extinction of our own species. On a probability curve the impacts in this report are usually at the very far right with a relatively low probability compared with other impacts, illustrated by (D) in Figure 2. Often they are so far out on the tail of the curve that they are not even included in studies. For each risk in this report the probability of an infinite impact is very low compared to the most likely outcome. Some studies even indicate that not all risks in this report can result in an infinite impact. But a significant number of **peer-reviewed** reports indicate that those impacts **not only can happen**, but that their **probability is increasing** due to **unsustainable trends**. The assumption for this report is that by creating a better understanding of our scientific knowledge regarding risks with a potentially infinite impact, we can inspire initiatives that can turn these risks into drivers for innovation. Not only could a better understanding of the unique magnitude of these risks help address the risks we face, it could also help to create a path towards more sustainable development. The group of global risks discussed in this report are so different from most of the challenges we face that they are hard to comprehend. But that is also why they can help us to build the collaboration we need and drive the development of further solutions that benefit both people and the planet. As noted above, none of the risks in this report is likely to result directly in an infinite impact, and some are probably even physically incapable of doing so. But all are so significant that they could reach a threshold impact able to create social and ecological instability that could trigger a process which could lead to an infinite impact. For several reasons the potentially infinite impacts of the risks in this report are not as well known as they should be. One reason is the way that extreme impacts are often masked by most of the theories and models used by governments and business today. For example, the probability of extreme impacts is often below what is included in studies and strategies. The tendency to exclude impacts below a probability of five percent is one reason for the **relative “invisibility”** of **infinite impacts**. The almost standard use of a 95% confidence interval is one reason why low-probability high-impact events are often ignored.6 Figure 2: Probability density function with tail highlighted [FIGURE 2 OMITTED] Climate change is a good example, where almost all of the focus is on the most likely scenarios and there are few studies that include the low-probability high-impact scenarios. In most reports about climate impacts, the impacts caused by warming beyond five or six degrees Celsius are even omitted from tables and graphs even though the IPCC’s own research indicates that the probability of these impacts are often between one and five percent, and sometimes even higher.7 Other aspects that contribute to this relative invisibility include the fact that extreme impacts are difficult to translate into monetary terms, they have a global scope, and they often require a time-horizon of a century or more. They cannot be understood simply by linear extrapolation of current trends, and they lack historical precedents. There is also the fact that the measures required to significantly reduce the probability of infinite impacts will be radical compared to a business-as-usual scenario with a focus on incremental changes. The exact probability of a specific impact is difficult or impossible to estimate.8 However, the important thing is to establish the current magnitude of the probabilities and compare them with the probabilities for such impacts we cannot accept. A failure to provide any estimate for these risks often results in strategies and priorities defined as though the probability of a totally unacceptable outcome is zero. An approximate number for a best estimate also makes it easier to understand that a great uncertainty means the actual probability can be both much higher and much lower than the best estimate. It should also be stressed that uncertainty is not a weakness in science; it always exists in scientific work. It is a systematic way of understanding the limitations of the methodology, data, etc.9 Uncertainty is not a reason to wait to take action if the impacts are serious. Increased uncertainty is something that risk experts, e.g. insurance experts and security policy experts, interpret as a signal for action. A contrasting challenge is that our cultural references to the threat of infinite impacts have been dominated throughout history by religious groups seeking to scare society without any scientific backing, often as a way to **discipline people** and implement unpopular measures. It should not have to be said, but this report is obviously **fundamentally different** as it focuses on **scientific evidence** from **peer-reviewed sources**. Infinite impact The concept infinite impact refers to two aspects in particular; the terminology is not meant to imply a literally infinite impact (with all the mathematical subtleties that would imply) but to serve as a reminder that these risks are of a different nature. Ethical These are impacts that threaten the very **survival of humanity and life on Earth** – and therefore can be seen as being infinitely negative from an **ethical perspective**. **No positive gain can outweigh even a small probability** for an infinite negative impact. Such risks require society to ensure that we eliminate these risks by reducing the impact below an infinite impact as a **top priority**, or at least do everything we can to reduce the probability of these risks. As some of these risks are impossible to eliminate today it is also important to discuss what probability can right now be accepted for risks with a possible infinite impact. Economic Infinite impacts are beyond what most traditional economic models today are able to cope with. The impacts are irreversible in the most fundamental way, so tools like cost-benefit assessment seldom make sense. To use discounting that makes infinite impacts (which could take place 100 years or more from now and affect all future generations) close to invisible in economic assessments, is another example of a challenge with current tools. So while tools like cost-benefit models and discounting can help us in some areas, they are seldom applicable in the context of infinite impacts. New tools are needed to guide the global economy in an age of potential infinite impacts. See chapter 2.2.2 for a more detailed iscussion. Roulette and Russian roulette When probability and normal risks are discussed the example of a casino and roulette is often used. You bet something, then spin the wheel and with a certain probability you win or lose. You can use different odds to discuss different kinds of risk taking. These kinds of thought experiment can be very useful, but when it comes to infinite risks these gaming analogies become problematic. For infinite impact a more appropriate analogy is probably Russian roulette. But instead of “normal” Russian roulette where you only bet your own life you are now also betting everyone you know and everyone you don’t know. Everyone alive will die if you lose. There will be no second chance for anyone as there will be no future generations; humanity will end with your loss. What probability would you accept for different sums of money if you played this version of Russian roulette? Most people would say that it is stupid and – no matter how low the probability is and no matter how big the potential win is – this kind of game should not be played, as it is unethical. Many would also say that no person should be allowed to make such a judgment, as those who are affected do not have a say. You could add that most of those who will lose from it cannot say anything as they are not born and will never exist if you lose. The difference between ordinary roulette and “allhumanity Russian roulette” is one way of illustrating the difference in nature between a “normal” risk that is reversible, and a risk with an infinite impact. An additional challenge in acknowledging the risks outlined in this report is that many of the traditional risks including wars and violence have decreased, even though it might not always looks that way in media.10 So a significant number of experts today spend a substantial amount of time trying to explain that much of what is discussed as dangerous trends might not be as dangerous as we think. For policy makers listening only to experts in traditional risk areas it is therefore easy to get the impression that global risks are becoming less of a problem. The **chain of events** that could result in **infinite impacts** in this report also **differ** from most of the traditional risks, as most of them are **not** triggered by **wilful acts**, but **accidents**/mistakes. Even the probabilities related to nuclear war in this report are to a large degree related to inadvertent escalation. As many of the tools to analyse and address risks have been developed to protect nations and states from attacks, risks involving accidents tend to get less attention. This report emphasises the need for an open and democratic process in addressing global challenges with potentially infinite impact. Hence, this is a scientifically based invitation to discuss how we as a global community can address what could be considered the greatest challenges of our time. The difficulty for individual scientists to communicate a scientific risk approach should however not be underestimated. Scientists who today talk about low-probability impacts, that are serious but still far from infinite, are often accused of pessimism and scaremongering, even if they do nothing but highlight scientific findings.11 To highlight infinite impacts with even lower probability can therefore be something that a scientist who cares about his/her reputation would want to avoid. In the media it is still common to contrast the most probable climate impact with the probability that nothing, or almost nothing, will happen. The fact that almost nothing could happen is not wrong in most cases, but it is unscientific and dangerous if different levels of probability are presented as equal. The tendency to compare the most probable climate impact with the possibility of a low or no impact also results in a situation where low-probability high-impact outcomes are often totally ignored. An honest and scientific approach is to, whenever possible, present the whole probability distribution and pay special attention to unacceptable outcomes. The fact that we have challenges that with some probability might be infinite and therefore fundamentally irreversible is difficult to comprehend, and **physiologically** they are something our brains are poorly equipped to respond to, according to evolutionary psychologists.12 It is hard for us as individuals to grasp that humanity for the first time in its history now has the capacity to create such catastrophic outcomes. Professor Marianne Frankenhaeuser, former head of the psychology division, Karolinska Institute, Stockholm, put it this way: “Part of the answer is to be found in psychological defence mechanisms. The nuclear threat is collectively denied, because to face it would force us to face some aspects of the world’s situation which we do not want to recognise.” 13 This **psychological** denial may be one reason why there is a tendency among some stakeholders to confuse “being optimistic” with denying what science is telling us, and ignoring parts of the probability curve.14 Ignoring the fact that there is strong scientific evidence for serious impacts in different areas, and focusing only on selected sources which suggest that the problem may not be so serious, is not optimistic. It is both unscientific and dangerous.15 A scientific approach requires us to base our decisions on the whole probability distribution. Whether it is possible to address the challenge or not is the area where optimism and pessimism can make people look at the same set of data and come to different conclusions. Two things are important to keep in mind: first, that there is always a probability distribution when it comes to risk; second, that there are two different kinds of impacts that are of interest for this report. The probability distribution can have different shapes but in simplified cases the shape tends to look like a slightly modified clock (remember figure 1). In the media it can sound as though experts argue whether an impact, for example a climate impact or a pandemic, will be dangerous or not. But what serious experts discuss is the probability of different oucomes. They can disagree on the shape of the curve or what curves should be studied, but not that a probability curve exists. With climate change this includes discussions about how sensitive the climate is, how much greenhouse gas will be emitted, and what impacts that different warmings will result in. Just as it is important not to ignore challenges with potentially infinite impacts, it is also important not to use them to scare people. Dramatic images and strong language are best avoided whenever possible, as this group of risks require sophisticated strategies that benefit from rational arguments. Throughout history we have seen too many examples when threats of danger have been damagingly used to undermine important values. The history of infinite impacts: The LA-602 document The understanding of infinite impacts is very recent compared with most of our institutions and laws. It is only 70 years ago that Edward Teller, one of the greatest physicists of his time, with his back-of-the-envelope calculations, produced results that differed drastically from all that had gone before. His calculations indicated that the explosion of a nuclear bomb – a creation of some of the brightest minds on the planet, including Teller himself – could result in a chain reaction so powerful that it would ignite the world’s atmosphere, thereby ending human life on Earth.16 Robert Oppenheimer, who led the Manhattan Project to develop the nuclear bomb, halted the project to see whether Teller’s calculations were correct.17 The resulting document, LA- 602: Ignition of the Atmosphere with Nuclear Bombs, concluded that Teller was wrong, But the sheer complexity drove them to end their assessment by writing that “further work on the subject [is] highly desirable”.18 The LA-602 document can be seen as the first scientific global risk report addressing a category of risks where the worst possible impact in all practical senses is infinite.19 Since the atomic bomb more challenges have emerged with potentially infinite impact. Allmost all of these new challenges are linked to the increased knowledge, economic and technical development that has brought so many benefits. For example, climate change is the result of the industrial revolution and development that was, and still is, based heavily on fossil fuel. The increased potential for global pandemics is the result of an integrated global economy where goods and services move quickly around the world, combined with rapid urbanisation and high population density. In parallel with the increased number of risks with possible infinite impact, our capacity to analyse and solve them has greatly increased too. Science and technology today provides us with knowledge and tools that can radically reduce the risks that historically have been behind major extinctions, such as pandemics and asteroids. Recent challenges like climate change, and emerging challenges like synthetic biology and nanotechnology, can to a large degree be addressed by smart use of new technologies, new lifestyles and institutional structures. It will be hard as it will require collaboration of a kind that we have not seen before. It will also require us to create systems that can deal with the problems before they occur. The fact that the same knowledge and tools can be both a problem and a solution is important to understand in order to avoid polarisation. Within a few decades, or even sooner, many of the tools that can help us solve the global challenges of today will come from fields likely to provide us with the most powerful instruments we have ever had – resulting in their own sets of challenges. Synthetic biology, nanotechnology and artificial intelligence (AI) are all rapidly evolving fields with great potential. They may help solve many of today’s main challenges or, if not guided in a benign direction, may result in catastrophic outcomes. The point of departure of this report is the fact that we now have the knowledge, economic resources and technological ability to reduce most of the greatest risks of our time. Conversely, the infinite impacts we face are almost all unintended results of human ingenuity. The reason we are in this situation is that we have made progress in many areas without addressing unintended low-probability high-impact consequences. Creating innovative and resilient systems rather than simply managing risk would let us focus more on opportunities. But the resilience needed require moving away from legacy systems is likely to be disruptive, so an open and transparent discussion is needed regarding the transformative solutions required. Figure 3: Probability density function with tail and threshold highlighted [FIGURE 3 OMITTED] 2.1 Report structure The first part of the report is an introduction where the global risks with potential infinite impact are introduced and defined. This part also includes the methodology for selecting these risks, and presents the twelve risks that meet this definition. Four goals of the report are also presented, under the headings “acknowledge”, “inspire”, “connect” and “deliver”. The second part is an overview of the twelve global risks and key events that illustrate some of the work around the world to address them. For each challenge five important factors that influence the probability or impact are also listed. The risks are divided into four different categories depending on their characteristics. “Current challenges” is the first category and includes the risks that currently threaten humanity due to our economic and technological development - extreme climate change, for example, which depends on how much greenhouse gas we emit. “Exogenic challenges” includes risks where the basic probability of an event is beyond human control, but where the probability and magnitude of the impact can be influenced - asteroid impacts, for example, where the asteroids’ paths are beyond human control but an impact can be moderated by either changing the direction of the asteroid or preparing for an impact. “Emerging challenges” includes areas where technological development and scientific assessment indicate that they could both be a very important contribution to human welfare and help reduce the risks associated with current challenges, but could also result in new infinite impacts.20 AI, nanotechnology and synthetic biology are examples. “Global policy challenge” is a different kind of risk. It is a probable threat arising from future global governance as it resorts to destructive policies, possibly in response to the other challenges listed above. The third part of the report discusses the relationship between the different risks. Action to reduce one risk can increase another, unless their possible links are understood. Many solutions are also able to address multiple risks, so there are significant benefits from understanding how one relates to others. Investigating these correlations could be a start, but correlation is a linear measure and non-linear techniques may be more helpful for assessing the aggregate risk. The fourth part is an overview, the first ever to our knowledge, of the uncertainties and probabilities of global risks with potentially infinite impacts. The numbers are only rough estimates and are meant to be a first step in a dialogue where methodologies are developed and estimates refined. The fifth part presents some of the most important underlying trends that influence the global challenges, which often build up slowly until they reach a threshold and very rapid changes ensue. The sixth and final part presents an overview of possible ways forward. 2.2 Goals Goal 1: Acknowledge That key stakeholders, influencing global challenges, acknowledge the existence of the category of risks that could result in infinite impact. They should also recognice that the list of risks that belong to this category should be revised as new technologies are developed and our knowledge increases. Regardless of the risks included, the category should be given special attention in all processes and decisions of relevance. The report also seeks to demonstrate to all key stakeholders that we have the capacity to reduce, or even eliminate, most of the risks in this category. Establish a category of risks with potentially infinite impact. Before anything significant can happen regarding global risks with potentially infinite impacts, their existence must be acknowledged. Rapid technological development and economic growth have delivered unprecedented material welfare to billions of people in a veritable tide of utopias.21 But we now face the possibility that even tools created with the best of intentions can have a darker side too, a side that may threaten human civilisation, and conceivably the continuation of human life. This is what all decision-makers need to recognise. Rather than succumbing to terror, we need to acknowledge that we can let the prospect inspire and drive us forward. Goal 2: Inspire That policy makers inspire action by explaining how the probabilities and impacts can be reduced and turned into opportunities. Concrete examples of initiatives should be communicated in different networks in order to create ripple effects, with the long-term goal that all key stakeholders should be inspired to turn these risks into opportunities for positive action. Show concrete action that is taking place today. This report seeks to show that it is not only possible to contribute to reducing these risks, but that it is perhaps the most important thing anyone can spend their time on. It does so by combining information about the risks with information about individuals and groups who has made a significant contribution by turning challenges into opportunities. By highlighting concrete examples the report hopes to inspire a new generation of leaders. Goal 3: Connect That leaders in different sectors connect with each other to encourage collaboration. A specific focus on financial and security policy where significant risks combine to demand action beyond the incremental is required. Support new meetings between interested stakeholders. The nature of these risks spans countries and continents; they require action by governments and politicians, but also by companies, academics, NGOs, and many other groups. The magnitude of the possible impacts requires not only leaders to act but above all new models for global cooperation and decision-making to ensure delivery. The need for political leadership is therefore crucial. Even with those risks where many groups are involved, such as climate change and pandemics, very few today address the possibility of infinite impact aspects. Even fewer groups address the links between the different risks. There is also a need to connect different levels of work, so that local, regional, national and international efforts can support each other when it comes to risks with potentially infinite impacts. Goal 4: Deliver That concrete strategies are developed that allow key stakeholders to identify, quantify and address global challenges as well as gather support for concrete steps towards a wellfunctioning global governance system. This would include tools and initiatives that can help identify, quantify and reduce risks with potentially infinite impacts. Identify and implement strategies and initiatives. Reports can acknowledge, inspire and connect, but only people can deliver actual results. The main focus of the report is to show that actual initiatives need to be taken that deliver actual results. Only when the probability of an infinite impact becomes acceptably low, very close to zero, and/or when the maximum impact is significantly reduced, should we talk about real progress. In order to deliver results it is important to remember that global governance to tackle these risks is the way we organise society in order to address our greatest challenges. It is not a question of establishing a “world government”, it is about the way we organise ourselves on all levels, from the local to the global. The report is a first step and should be seen as an invitation to all responsible parties that can affect the probability and impact of risks with potentially infinite impacts. But its success will ultimately be measured only on how it contributes to concrete results. 2.3 Global challenges and infinite impact This chapter first introduces the concept of infinite impact. It then describes the methodology used to identify challenges with an infinite impact. It then presents risks with potentially infinite impact that the methodology results in. 2.3.1 Definition of infinite impact The specific criterion for including a risk in this report is that well-sourced science shows the challenge can have the following consequences: 22 1. Infinite impact: When civilisation collapses to a state of great suffering and does not recover, or a situation where all human life ends. The existence of such threats is well attested by science.23 2. Infinite impact threshold – an impact that can trigger a chain of events that could result first in a civilisation collapse, and then later result in an infinite impact. Such thresholds are especially important to recognise in a complex and interconnected society where resilience is decreasing.24 A collapse of civilisation is defined as a drastic decrease in human population size and political/economic/social complexity, globally for an extended time.25 The above definition means the list of challenges is not static. When new challenges emerge, or current ones fade away, the list will change. An additional criterion for including risks in this report is “human influence”. Only risks where humans can influence either the probability, the impact, or both, are included. For most risks both impact and probability can be affected, for example with nuclear war, where the number/size of weapons influences the impact and tensions between countries affects the probability. Other risks, such as a supervolcano, are included as it is possible to affect the impact through various mitigation methods, even if we currently cannot affect the probability. Risks that are susceptible to human influence are indirectly linked, because efforts to address one of them may increase or decrease the likelihood of another. 2.3.2 Why use “infinite impact” as a concept? The concept of infinity was chosen as it reflects many of the challenges, especially in economic theory, to addressing these risks as well as the need to question much of our current way of thinking. The concept of a category of risks based on their extreme impact is meant to provide a tool to distinguish one particular kind of risk from others. The benefit of this new concept should be assessed based on two things. First, does the category exist, and second, is the concept helpful in addressing these risks? The report has found ample evidence that there are risks with an impact that can end human civilisation and even all human life. The report further concludes that a new category of risk is not only meaningful but also timely. We live in a society where global risks with potentially infinite impacts increase in both number and probability according to multiple studies. Looking ahead, many emerging technologies which will certainly provide beneficial results, might also result in an increased probability of infinite impacts.26 Over the last few years a greater understanding of low probability or unknown probability events has helped more people to understand the importance of looking beyond the most probable scenarios. Concepts like “black swans” and “perfect storms” are now part of mainstream policy and business language.27 Greater understanding of the technology and science of complex systems has also resulted in a new understanding of potentially disruptive events. Humans now have such an impact on the planet that the term “the anthropocene” is being used, even by mainstream media like The Economist.28 The term was introduced in the 90s by the Nobel Prize winner Paul Crutzen to describe how humans are now the dominant force changing the Earth’s ecosystems.29 The idea to establish a well defined category of risks that focus on risks with a potentially infinite impact that can be used as a practical tool by policy makers is partly inspired by Nick Bostrom’s philosophical work and his introduction of a risk taxonomy that includes an academic category called “existential risks”.30 Introducing a category with risks that have a potentially infinite impact is not meant to be a mathematical definition; infinity is a thorny mathematical concept and nothing in reality can be infinite.31 It is meant to illustrate a singularity, when humanity is threatened, when many of the tools used to approach most challenges today become problematic, meaningless, or even counterproductive. The concept of an infinite impact highlights a unique situation where humanity itself is threatened and the very idea of value and price collapses from a human perspective, as the price of the last humans also can be seen to be infinite. This is not to say that those traditional tools cannot still be useful, but with infinite impacts we need to add an additional set of analytical tools. Life Value The following estimates have been applied to the value of life in the US. The estimates are either for one year of additional life or for the statistical value of a single life. – $50,000 per year of quality life (international standard most private and government-run health insurance plans worldwide use to determine whether to cover a new medical procedure) – $129,000 per year of quality life (based on analysis of kidney dialysis procedures by Stefanos Zenios and colleagues at Stanford Graduate School of Business) – $7.4 million (Environmental Protection Agency) – $7.9 million (Food and Drug Administration) – $6 million (Transportation Department) – $28 million (Richard Posner based on the willingness to pay for avoiding a plane crash) Source: Wikipedia: Value of life http://en.wikipedia.org/wiki/Value\_of\_life US EPA: Frequently Asked Questions on Mortality Risk Valuation http://yosemite.epa.gov/EE%5Cepa%5Ceed.nsf/webpages/MortalityRiskValuation.html Posner, Richard A. Catastrophe: risk and response. Oxford University Press, 2004 Some of the risks, including nuclear war, climate change and pandemics, are often included in current risk overviews, but in many cases their possible infinite impacts are excluded. The impacts which are included are in most cases still very serious, but only the more probable parts of the probability distributions are included, and the last part of the long tail – where the infinite impact is found – is excluded.32 Most risk reports do not differentiate between challenges with a limited impact and those with a potential for infinite impact. This is dangerous, as it can mean resources are spent in ways that increase the probability of an infinite impact. Ethical aspects of infinite impact The basic ethical aspect of infinite impact is this: a very small group alive today can take decisions that will fundamentally affect all future generations. “All future generations” is not a concept that is often discussed, and for good reason. All through human history we have had no tools with a measurable global impact for more than a few generations. Only in the last few decades has our potential impact reached a level where all future generations can be affected, for the simple reason that we now have the technological capacity to end human civilisation. If we count human history from the time when we began to practice settled agriculture, that gives us about 12,000 years.33 If we make a moderate assumption that humanity will live for at least 50 million more years34 our 12,000-year history so far represents 1/4200, or 0.024%, of our potential history. So our generation has the option of risking everything and annulling 99.976% of our potential history. Comparing 0.024% with the days of a person living to 100 years from the day of conception, this would equal less than nine days and is the first stage of human embryogenesis, the germinal stage.35 Two additional arguments to treat potentially infinite impacts as a separate category are: 36 1. **An approach to infinite impacts cannot be one of trial-and-error**, because there is no opportunity to learn from errors. The reactive approach – see what happens, limit damage, and learn from experience – is unworkable. Instead society must be proactive. This requires foresight to foresee new types of threat and willingness to take decisive preventative action and to bear the costs (**moral** and economic) of such actions. 2. We cannot necessarily rely on the institutions, morality, social attitudes or national security policies that developed from our experience of other sorts of risk. Infinite impacts are in a different category. Institutions and individuals may find it hard to take these risks seriously simply because they lie outside our experience. Our collective fear-response will probably be ill-calibrated to the magnitude of threat. Economic aspects of infinite impact and discounting In today’s society a monetary value is sometimes ascribed to human life. Some experts use this method to estimate risk by assigning a monetary value to human extinction.37 We have to remember that the monetary values placed on a human life in most cases are not meant to suggest that we have actually assigned a specific value to a life. Assigning a value to a human life is a tool used in a society with a limited supply of resources or infrastructure (ambulances, perhaps) or skills. In such a society it is impossible to save every life, so some trade-off must be made.38 The US Environmental Protection Agency explains its use like this: “The EPA does not place a dollar value on individual lives. Rather, when conducting a benefit-cost analysis of new environmental policies, the Agency uses estimates of how much people are willing to pay for small reductions in their risks of dying from adverse health conditions that may be caused by environmental pollution.” 39 The fact that monetary values for human lives can help to define priorities when it comes to smaller risks does not mean that they are suitable for quite different uses. Applying a monetary value to the whole human race makes little sense to most people, and from an economic perspective it makes no sense. Money helps us to prioritise, but with no humans there would be no economy and no need for priorities. Ignoring, or discounting, future generations is actually the only way to avoid astronomical numbers for impacts that may seriously affect every generation to come. In Catastrophe: Risk and Response, Richard Posner provides a cost estimate, based on the assumption that a human life is worth $50,000, resulting in a $300 tn cost for the whole of humanity, assuming a population of six billion. He then doubles the population number to include the value of all future generations, ending up with $600 tn, while acknowledging that “without discounting, the present value of the benefits of risk-avoidance measures would often approach infinity for the type of catastrophic risk with which this book is concerned.” 40 Discounting for risks that include the possibility of an infinite impact differs from risk discounting for less serious impacts. For example the Stern Review41 prompted a discussion between its chief author, Nicholas Stern, and William Nordhaus,42 each of whom argued for different discount levels using different arguments. But neither discussed a possible infinite climate impact. An overview of the discussion by David Evans of Oxford Brookes University highlighted some of the differing assumptions.43 Two things make infinite impacts special from a discounting perspective. First, there is no way that future generations can compensate for the impact, as they will not exist. Second, the impact is something that is beyond an individual preference, as society will no longer exist. Discounting is undertaken to allocate resources in the most productive way. In cases that do not include infinite impacts, discounting “reflects the fact that there are many high-yield investments that would improve the quality of life for future generations. The discount rate should be set so that our investable funds are devoted to the most productive uses.” 44 When there is a potentially infinite impact, the focus is no longer on what investments have the best rate of return, it is about avoiding the ultimate end. While many economists shy away from infinite impacts, those exploring the potentially extreme impacts of global challenges often assume infinite numbers to make their point. Nordhaus for example writes that “the sum of undiscounted anxieties would be infinite (i.e. equal to 1 + 1 +1 + … = ∞). In this situation, most of us would dissolve in a sea of anxiety about all the things that could go wrong for distant generations from asteroids, wars, out-of-control robots, fat tails, smart dust and other disasters.” 45 It is interesting that Nordhaus himself provides very good graphs that show why the most important factor when determining actions is a possible threshold (see below Figure 4 and 5). Nordhaus was discussing climate change, but the role of thresholds is similar for most infinite impacts. The first figure is based on traditional economic approaches which assume that Nature has no thresholds; the second graph illustrates what happens with the curve when a threshold exists. As Nordhaus also notes, it is hard to establish thresholds, but if they are significant all other assumptions become secondary. The challenge that Nordhaus does not address, and which is important especially with climate change, is that thresholds become invisible in economic calculations if they occur far into the future, even if it is **current actions** that unbalance the system and eventually push it over the threshold.46 Note that these dramatic illustrations rest on assumptions that the thresholds are still relatively benign, not moving us beyond tipping points which result in an accelerated release of methane that could result in a temperature increase of more than 8 °C, possibly producing infinite impacts.47 Calculating illustrative numbers By including the welfare of future generations, something that is important when their very existence is threatened, economic discounting becomes difficult. In this chapter, some illustrative numbers are provided to indicate the order of magnitude of the values that calculations provide when traditional calculations also include future generations. These illustrative calculations are only illustrative as the timespans that must be used make all traditional assumptions questionable to say the least. Still, as an indicator for why infinite impact might be a good approximation they might help. As a species that can manipulate our environment it could be argued that the time the human race will be around, if we do not kill ourselves, can be estimated to be between 1-10 million years – the typical time period for the biological evolution of a successful species48 – and one billion years, the inhabitable time of Earth.49 [FIGURE 4 OMITTED] [FIGURE 5 OMITTED] If we assume – 50 million years for the future of humanity as our reference, – an average life expectancy of 100 years50, and – a global population of 6 billion people51 – all conservative estimate – , we have half a million generations ahead of us with a total of 3 quadrillion individuals. Assuming a value of $50,000 per life, the cost of losing them would then be $1.5 ×1020, or $150 quintillion. This is a very low estimate, and Posner suggests that maybe the cost of a life should be “written up $28 million” for catastrophic risks52. Posner’s calculations where only one future generation is included result in a cost of $336 quadrillion. If we include all future generations with the same value, $28 million, the result is a total cost of $86 sextillion, or $86 × 1021. This $86 sextillion is obviously a very rough number (using one billion years instead of 50 million would for example require us to multiply the results by 20), but again it is the magnitude that is interesting. As a reference there are about 1011 to 1012 stars in our galaxy, and perhaps something like the same number of galaxies. With this simple calculation you get 1022 to 1024, or 10 to 1,000 sextillion, stars in the universe to put the cost of infinite impacts when including future generations in perspective.53 These numbers can be multiplied many times if a more philosophical and technology-optimistic scenario is assumed for how many lives we should include in future generations. The following quote is from an article by Nick Bostrom in Global Policy Journal: “However, the relevant figure is not how many people could live on Earth but how many descendants we could have in total. One lower bound of the number of biological human life-years in the future accessible universe (based on current cosmological estimates) is 1034 years. Another estimate, which assumes that future minds will be mainly implemented in computational hardware instead of biological neuronal wetware, produces a lower bound of 1054 human-brain-emulation subjective life-years.” 54 Likewise the value of a life, $28 million, a value that is based on an assessment of how individuals chose when it comes to flying, can be seen as much too small. This value is based on how much we value our own lives on the margin, and it is reasonable to assume that the value would be higher than only a multiplication of our own value if we also considered the risk of losing our family, everyone we know, as well as everyone else on the planet. In the same way as the cost increases when a certain product is in short supply, the cost of the last humans could be assumed to be very high, if not infinite. Obviously, the very idea to put a price on the survival of humanity can be questioned for good reasons, but if we still want to use a number, $28 million per life should at least be considered as a significant underestimation. For those that are reluctant or unable to use infinity in calculations and are in need of a number for their formulas, $86 sextillion could be a good initial start for the cost of infinite impacts. But it is important to note that this number might be orders of magnitude smaller than an estimate which actually took into account a more correct estimation of the number of people that should be included in future generations as well as the price that should be assigned to the loss of the last humans. 2.3.3 Infinite impact threshold (IIT) As we address very complex systems, such as human civilisation and global ecosystems, a concept as important as infinite impact in this report is that of infinity impact threshold. This is the impact level that can trigger a chain of events that results in the end of human civilisation. The infinite impact threshold (IIT) concept represents the idea that long before an actual infinite impact is reached there is a tipping point where it (with some probability) is no longer possible to reverse events. So instead of focusing only on the ultimate impact it is important to estimate what level of impact the infinity threshold entails. The IIT is defined as an impact that can trigger a chain of events that could result first in a civilisation collapse, and then later result in an infinite impact. Such thresholds are especially important to recognise in a complex and interconnected society where resilience is decreasing. Social and ecological systems are complex, and in most complex systems there are thresholds where positive feedback loops become self-reinforcing. In a system where resilience is too low, feedback loops can result in a total system collapse. These thresholds are very difficult to estimate and in most cases it is possible only to estimate their order of magnitude. As David Orrell and Patrick McSharry wrote in A Systems Approach to Forecasting: “Complex systems have emergent properties, qualities that cannot be predicted in advance from knowledge of systems components alone”. According to complexity scientist Stephen Wolfram’s principle of computational irreducibility, the only way to predict the evolution of such a system is to run the system itself: “There is no simple set of equations that can look into its future.” 55 Orrell and McSharry also noted that “in orthodox economics, the reductionist approach means that the economy is seen as consisting of individual, independent agents who act to maximise their own utility. It assumes that prices are driven to a state of near-equilibrium by the ‘invisible hand’ of the economy. Deviations from this state are assumed to be random and independent, so the price fluctuations are often modelled using the normal distribution or other distributions with thin tails and finite variance.” The drawbacks of an approach using the normal distribution, or other distributions with thin tails and finite variance, become obvious when the unexpected happens as in the recent credit crunch, when existing models totally failed to capture the true risks of the economy. As an employee of Lehman Brothers put it on August 11, 2007: “Events that models predicted would happen only once in 10,000 years happened every day for three days.” 56 [FIGURE 6 OMITTED] The exact level for an infinite impact threshold should not be the focus, but rather the fact that such thresholds exists and that an order of magnitude should be estimated.57 During the process of writing the report, experts suggested that a relatively quick death of two billion people could be used as a tentative number until more research is available.58 With current trends undermining ecological and social resilience it should be noted that the threshold level is likely to become lower as time progress. 2.3.4 Global F-N curves and ALARP In the context of global risks with potentially infinite impact, the possibility of establishing global F-N curves is worth exploring. One of the most common and flexible frameworks used for risk criteria divides risks into three bands: 59 1. Upper: an unacceptable/ intolerable region, where risks are intolerable except in extraordinary circumstances and risk reduction measures are essential. 2. Middle: an ALARP (“as low as reasonably practicable”) region, where risk reduction measures are desirable but may not be implemented if their cost is disproportionate to the benefit achieved. 3. Lower: a broadly acceptable/ negligible region, where no further risk reduction measures are needed. The bands are expressed by F-N curves. When the frequency of events which cause at least N fatalities is plotted against the number N on log–log scales, the result is called an F-N curve.60 If the frequency scale is replaced by annual probability, then the resultant curve is called an f-N curve. The concept for the middle band when using F-N curves is ALARP. It is a term often used in the area of safety-critical and safety-involved systems.62 The ALARP principle is that the residual risk should be as low as reasonably practicable. The upper band, the unacceptable/ intolerable region, is usually the area above the ALARP area (see figure 8) By using F-N curves it is also possible to establish absolute impact levels that are never acceptable, regardless of probability (Figure 7. Based on an actual F-n Curve showing an absolute impact level that is defined as unacceptable). This has been done in some cases for local projects. The infinite threshold could be used to create an impact limit on global F-N curves used for global challenges in the future. Such an approach would help governments, companies and researchers when they develop new technical solutions and when investing in resilience. Instead of reducing risk, such an approach encourages the building of systems which cannot have negative impacts above a certain level. Pros – Clearly shows relationship between frequency and size of accident – Allows judgement on relative importance of different sizes of accident – Slope steeper than -1 provides explicit consideration of multiple fatality aversion and favours concepts with lower potential for large fatality events – Allows company to manage overall risk exposure from portfolio of all existing and future facilities Cons – Cumulative expression makes it difficult to interpret, especially by non-risk specialists – Can be awkard to derive – May be difficult to use if criterion is exceeded in one area but otherwise is well below – Much debate about criterion lines Figure 7: Example of F-n curve showing different levels of risk 61 Figure 9: Pros and cons of F-N curves 63 46 Global Challenges – Twelve risks that threaten human civilisation – The case for a new category of risks 2.3 Global challenges and infinite impact practical guidance that can provide defined group of risks 2.3.5 A name for a clearly 10 100 1000 10000 10 10 10 10 10 10 10 10-2 -3 -4 -5 -6 -7 -8 -9 Number of Fatalities (N) Frequency (F) of Accidents with N or More Fatalities (Per Year) ALARP region Unacceptable Acceptable Today no established methodology exists that provides a constantly updated list of risks that threaten human civilisation, or even all human life. Given that such a category can help society to better understand and act to avoid such risks, and better understand the relation between these risks, it can be argued that a name for this category would be helpful.65 To name something that refers to the end of humanity is in itself a challenge, as the very idea is so far from our usual references and to many the intuitive feeling will be to dismiss any such thing. The concept used in this report is “infinity”. The reson for this is that many of the challenges relate to discussed. In one way the name is not very important so long as people understand the impacts and risks associated with it. Still, a name is symbolic and can either help or make it more difficult to get support to establish the new category. The work to establish a list of risks with infinite impact evolved from “existential risk”, the philosophical concept that inspired much of the work to establish a clearly defined group of risks. The reason for not using the concept “existential risk and impact” for this category, beside the fact that existential impact is also used in academic contexts to refer to a personal impact, is that the infinite category is a smaller subset of “existential risk” and this new category is meant to be used as a tool, not a scientific concept. Not only should the impacts in the category potentially result in the end of all human life, it should be possible to affect the probability and/or impact of that risk. There must also exist an agreed methodology, such as the one suggested in this report, that decides what risks belong and not belong on the list. Another concept that the category relates to is “global catastrophic risk” as it is one of the most used concepts among academics interested in infinite impacts. However it is vague enough to be used to refer to impacts from a few thousand deaths to the end of human civilisation. Already in use but not clearly defined, it includes both the academic concept existential risks and the category of risks with infinite impacts. macroeconomics and its challenges in relation to the kind of impacts that the risks in this report focus on. Further, the name clearly highlights the unique nature without any normative judgements. Still, infinity is an abstract concept and it might not be best communicate the unique group of risks that it covers to all stakeholders. In the same way as it can be hard to use singularity to describe a black hole, it can be difficult to use infinity to describe a certain risk. If people can accept that it is only from a specific perspective that the infinity concept is relevant it could be used beyond the areas of macroeconomics. Two other concepts that also have been considered during the process of writing this report are “xrisks” and “human risk of ruin”. Xrisk has the advantage, and disadvantage, of not really saying anything at all about the risk. The positive aspect is that the name can be associated with the general concept of extinction and the philosophical concept of existential risk as both have the letter x in them. The disadvantage is the x often represents the unknown and can therefore relate to any risk. There is nothing in the name that directly relates to the kind of impacts that the category covers, so it is easy to interpret the term as just unknown risks. Human risk of ruin has the advantage of having a direct link to a concept, risk of ruin, that relates to a very specific state where all is lost. Risk of ruin is a concept in use in gambling, insurance, and finance that can all give very important contributions to the work with this new category of risk. The resemblance to an existing concept that is well established could be both a strength and a liability. Below is an overview of the process when different names were Figure 8: Example of F-n curve showing an absolute impact level that is defined as unacceptable/ infinite. i.e no level of probability is acceptable above a certain level of impact, in this case 1000 dead 64 Global Challenges – Twelve risks that threaten human civilisation – The case for a new category of risks 47 2.3 Global challenges and infinite impact 3. 2. 1. 9. Unacceptable risks in different combinations, e.g. unacceptable global risks – This is probably not appropriate for two main reasons. First, it is a normative statement and the category aims to be scientific; whether these risks are unacceptable or not is up to the citizens of the world to decide. Second, the idea of risk is that it is a combination of probability times impact. If a risk is unacceptable is therefore also usually related to how easy it is to avoid. Even if a risk is small, due to relatively low probability and relatively low impact, but is very easy to address, it can be seen as unacceptable, in the same way a large risk can be seen as acceptable if it would require significant resources to reduce. There will not be a perfect concept and the question is what concept can find the best balance between being easy to understand, acceptable where policy decisions needs to be made and also acceptable for all key groups that are relevant for work in these area. During the process to find a name for this category inspiration has been found in the process when new concepts have been introduced; from irrational numbers and genocide to sustainable development and the Human Development Index. So far “infinite risk” can be seen as the least bad concept in some areas and “xrisks” and “human risk of ruin” the least bad in others. The purpose of this report is to establish a methodology to identify a very specific group of risks as well as continue to a process where these risks will be addressed in a systematic and appropriate way. The issue of naming this group of risks will be left to others. The important is that the category gets the attention it deserves. The three concepts are very different. Global catastrophic risk is possibly the most used concept in contexts where infinite impacts are included, but it is without any clear definition. Existential risk is an academic concept used by a much smaller group and with particular focus on future technologies. The category in this report is a tool to help decision makers develop strategies that help reduce the probability that humanity will end when it can be avoided. The relation between the three concepts can be illustrated with three circles. The large circle (1) represents global catastrophic risks, the middle one (2) existential risks and the small circle (3) the list of twelve risks in this report, i.e. risks where there are peer reviewed academic studies that estimate the probability of an infinite impact and where there are known ways to reduce the risk. A list that could be called infinite risks, xrisks, or human risk of ruin. Other concepts that are related to infinite impacts that could potentially be used to describe the same category if the above suggestions are not seen as acceptable concepts are presented below, together with the main reason why these concepts were not chosen for this report. 1. Risk of ruin – is a concept in gambling, insurance and finance relating to the likelihood of losing all one’s capital or affecting one’s bankroll beyond the point of recovery. It is used to describe individual companies rather than systems.66 2. Extinction risk – is used in biology for any species that is threatened. The concept is also used in memory/cognition research. It is a very dramatic term, to be used with care. These factors make it probably unsuitable for use by stakeholders accustomed to traditional risk assessment. 3. Astronomical risk – is seldom used scientifically, but when it is used it is often used for asteroids and is probably best reserved for them.67 4. Apocalyptic risk – could have been suitable, as the original meaning is apocálypsis, from the Greek ἀπό and καλύπτω meaning ‘un-covering’. It is sometime used, but in a more general sense, to mean significant risks.68 But through history and today it is mainly used for a religious end of time scenario. Its strong links to unscientific doom-mongers make it probably unsuitable for a scientific concept. 5. End-of-the-world risk - belongs to the irrational doomsday narratives and so is probably unsuitable for scientific risk assessments. 6. Extreme risk – is vague enough to describe anything beyond the normal, so it is probably unsuitable for risk assessments of this magnitude. 7. Unique risk – is even vaguer, as every risk is unique in some way. Probably best avoided in risk assessments. 8. Collapse risk – is based on Jared Diamond’s thinking.69 There are many different kinds of collapse and only a few result in infinite impact. 48 Global Challenges – Twelve risks that threaten human civilisation – The case for a new category of risks 2.3 Global challenges and infinite impact Estimations of impact Only literature where there is some estimation of impact that indicates the possibility of an infinite impact is included. Leading organisations’ priorities In order to increase the probability of covering all relevant risks an overview of leading organisations' work was conducted. This list was then compared with the initial list and subjected to the same filter regarding the possibility to affect the probability or impact. Possibility of addressing the risk Possibility of addressing the risk: From the risks gathered from literature and organisations, only those where the probability or impact can be affected by human actions are included. Expert review Qualitative assessment: Expert review in order to increase the probability of covering all relevant global risks. List of risks Result: List of risks with potentially infinite impacts. Relevant literature Identification of credible sources: search relevant literature in academic literature included in World of Knowledge and Google Scholar. 1 2 3 4 5 6 This chapter presents the methodology used to identify global risks with potentially infinite impact. Methodology overview In order to establish a list of global risks with potentially infinite impact a methodological triangulation was used, consisting of: – A quantitative assessment of relevant literature. – A strategic selection of relevant organisations and their priorities. – A qualitative assessment with the help of expert workshops. 2.4 Methodology 70 Global Challenges – Twelve risks that threaten human civilisation – The case for a new category of risks 49 2.4 Methodology The scientific review of literature was led by Seth Baum, Executive Director of the Global Catastrophic Risk Institute72 and research scientist at the Center for Research on Environmental Decisions, Columbia University.73 The methodology for including global risks with a potentially infinite impact is based on a scientific review of key literature, with focus on peer-reviewed academic journals, using keyword search of both World of Knowledge74 and Google Scholar75 combined with existing literature overviews in the area of global challenges. This also included a snowball methodology where references in the leading studies and books were used to identify other scientific studies and books. In order to select words for a literature search to identify infinite impacts, a process was established to identify words in the scientific literature connected to global challenges with potentially infinite impacts. Some words generate a lot of misses, i.e. publications that use the term but are not the focus of this report. For example “existential risk” is used in business; “human extinction” is used in memory/cognition. Some search terms produced relatively few hits. For example “global catastrophic risk” is not used much. Other words are only used by people within a specific research community: few use “existential risk” in our sense unless they are using Nick Bostrom’s work. The term “global catastrophe” was identified as a phrase that referred almost exclusively to extremely negative impacts on humans, by a diversity of researchers, not just people in one research community. A list of 178 relevant books and reports was established based on what other studies have referred to, and/or which are seen as landmark studies by groups interviewed during the process. They were selected for a closer examination regarding the challenges they include.76 The full bibliography, even with its focus on publications of general interest, is still rather long. So it is helpful to have a shorter list focused on the highlights; the most important publications based on how often they are quoted, how wellspread the content (methodology, lists, etc.) is and how often key organisations use them. The publications included must meet at least one of the following criteria: – Historical significance. This includes being the first publication to introduce certain key concepts, or other early discussions of global challenges. Publications of historical significance are important for showing the intellectual history of global challenges. Understanding how the state of the art research got to where it is today can also help us understand where it might go in the future. – Influential in developing the field. This includes publications that are highly cited77 and those that have motivated significant additional research. They are not necessarily the first publications to introduce the concepts they discuss, but for whatever reason they will have proved important in advancing research. – State of the art. This includes publications developing new concepts at the forefront of global challenges research as well as those providing the best discussions of important established concepts. Reading these publications would bring a researcher up to speed with current research on global challenges. So they are important for the quality of their ideas. – Covers multiple global challenges (at least two). Publications that discuss a variety of global challenges are of particular importance because they aid in identifying and comparing the various challenges. This process is essential for research on global risks to identify boundaries and research priorities. In order to identify which global challenges are most commonly discussed, key surveys were identified and coded. First, a list of publications that survey at least three global challenges was compiled, and they were then scanned to find which challenges they discussed. The publications that survey many global challenges were identified from the full bibliography. Publications from both the academic and popular literature were considered. Emphasis was placed on publications of repute or other significance.78 To qualify as a survey of global challenges, the publication had to provide an explicit list of challenges or to be of sufficient length and breadth for it to discuss a variety of challenges. Many of the publications are books or book-length collections of articles published in book form or as special issues of scholarly journals. Some individual articles were also included because they discussed a significant breadth of challenges. A total of 40 global challenge survey publications were identified. For authors with multiple entries (Bostrom with three and WEF with ten) each challenge was counted only once to avoid bias. review of key literature 71 2.4.1 A scientific 50 Global Challenges – Twelve risks that threaten human civilisation – The case for a new category of risks 2.4 Methodology 0 5 10 15 20 25 Climate Change Nuclear War Pandemic Biodiversity loss Asteroid / Comet / Meteor Volcano Genetic Engineering High Energy Physics Nanotech Resource Depletion Artificial Intelligence Chemical Pollution Ecological Catastrophe Biogeochem Government Failure Poverty System Failure Astronomic Explosion LULCC Biological Weapons Chemical Weapons Extraterrestrial Reject Procreation Computer Failure EM Pulse New Technology Ozone Depletion Dysgenics Ocean Acidification Interstellar Cloud Atmosphere Aerosols Phase Transition Simulation Unknown 21 18 17 15 14 14 13 13 13 13 11 11 11 8 8 8 8 7 7 5 5 5 5 4 4 4 4 3 3 2 1 1 1 1 In terms of authorship and audience, there are 17 academic publications, 9 popular publications, 1 government report, 3 publications written by academics for popular audiences. In terms of format, there are 15 books, 5 edited collections, 7 articles, 3 of miscellaneous format. Of the 40 publications identified, 22 were available at the time of coding. In addition, 10 Global Risks Reports from the World Economic Forum were coded and then gathered under one heading: “WEF Global Risk Report 2005-2014”. A list of 34 global challenges was developed based on the challenges mentioned in the publications. A spreadsheet containing the challenges and the publications was created to record mentions of specific challenges in each publication to be coded. Then each publication was scanned in its entirety for mentions of global challenges. Scanning by this method was necessary because many of the publications did not contain explicit lists of global challenges, and the ones that did often mentioned additional challenges separately from their lists. So it was not required that a global challenge be mentioned in a list for it to be counted – it only had to be mentioned somewhere in the publication as a challenge. Assessing whether a particular portion of text counts as a global challenge and which category it fits in sometimes requires some interpretation. This is inevitable for most types of textual analysis, or, more generally, for the coding of qualitative data. The need for interpretation in this coding was heightened by the fact that the publications often were not written with the purpose of surveying the breadth of global challenges, and even the publications that were intended as surveys did not use consistent definitions of global challenges. The coding presented here erred on the side of greater inclusivity: if a portion of text was in the vicinity of a global challenge, then it was coded as one. For example, some publications discussed risks associated with nuclear weapons in a general sense without specifically mentioning the possibility of large-scale nuclear war. These discussions were coded as mentions of nuclear war, even though they could also refer to single usages of nuclear weapons that would not rate as a global challenge. This more inclusive approach is warranted because many of the publications were not focused exclusively on global challenges. If they were focused on them, it is likely that they would have included these risks in their global challenge form (e.g., nuclear war), given that they were already discussing something related (e.g., nuclear weapons). Below are the results from the overview of the surveys. Figure 9: Number of times global challenges are included in surveys of global challenges Global Challenges – Twelve risks that threaten human civilisation – The case for a new category of risks 51 2.4 Methodology Climate Change Nuclear War Pandemic Biodiversity loss Asteroid / Comet / Meteor Volcano Genetic Engineering High Energy Physics Nanotech Resource Depletion Artificial Intelligence Chemical Pollution Ecological Catastrophe 21 18 17 15 14 14 13 13 13 13 11 11 11 0 25 20 15 10 5 dung beetle star trek zinc oxalate human extinction 0 200 400 600 800 1000 It should be noted that the literature that includes multiple global challenges with potentially infinite impact is very small, given the fact that it is about the survival of the human race. Experts in the field of global challenges, like Nick Bostrom, have urged policymakers and donors to focus more on the global challenges with infinite impacts and have used dramatic rhetoric to illustrate how little research is being done on them compared with other areas. However, it is important to note that many more studies exist that focus on individual global risks, but often without including low-probability high-impact outcomes.80 How much work actually exists on human extinction infinite impact is therefore difficult to assess. The list of risks found in the scientific literature was checked against a review of what challenges key organisations working on global challenges include in their material and on their webpages. This was done to ensure that no important risk was excluded from the list. The coding of key organisations paralleled the coding of key survey publications. Organisations were identified via the global catastrophic risk organisation directory published by the Global Catastrophic Risk Institute.82 They were selected from the directory if they worked on a variety of global challenges – at least three, and ideally more. The reason for focusing on those that work on multiple challenges is to understand which challenges they consider important and why. In contrast, organisations that focus on only one or two challenges may not Figure 10: The global challenges included ten times or more in surveys of global challenges on global challenges 81 organisations working 2.4.2 A review of Figure 11: Number of academic papers on various topics (listed in Scopus, August 2012) From the paper “Existential Risk Prevention as Global Priority” 79 52 Global Challenges – Twelve risks that threaten human civilisation – The case for a new category of risks 2.4 Methodology Climate Change Nuclear War Pandemic Resource Depletion Biological Weapons Computer Failure Government Failure Nanotech Chemical Weapons Artificial Intelligence Genetic Engineering System Failure Biodiversity loss Ecological Failure Poverty Volcano Asteroid / Comet / Meteor Astronomic Explosion Biogeochem Chemical Pollution Extraterrestrial High Energy Physics New Technology Ozone Depletion Atmospheric Aerosols Dysgenics EM Pulse Interstellar Cloud LULCC Ocean Acidification Phase Transition Reject Procreation Simulation Unknown 13 13 12 9 8 7 7 7 6 5 4 4 2 2 2 2 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 4 8 12 2 6 10 14 be able to adjust their focus according to which challenges they consider the most important. The organisation coding used the same coding scheme developed for coding survey publications. References to specific global challenges were obtained from organisations’ websites. Many have web pages which list the topics they work on. Where possible, references to global challenges were pulled from these pages. Additional references to these challenges were identified by browsing other web pages, including recent publications. While it is possible that some of these organisations have worked on global challenges not mentioned on the web pages that were examined, overall the main challenges that they have worked on have probably been identified and coded. So the results should give a reasonably accurate picture of what global challenges these organisations are working on. Organisations working with global challenges were initially selected on the basis of the literature overview. A snowball sampling was conducted based on the list of organisations identified, according to whether they claimed to work on global challenges and/or their web page contained information about “existential risk”, “global catastrophic risk”,“human extinction” or “greatest global challenges”. Cross-references between organisations and input during the workshops were also used to identify organisations. An initial list of 180 organisations which work with global challenges was established. Based on the production of relevant literature, which other organisations referred to the organisation, and/or are seen as influential by groups interviewed during the process, a short-list of organisations were selected for a closer examination regarding the challenges they work with. Then those working with multiple challenges were selected, resulting in a list of 19 organisations.83 Below is the overview of the results from the overview of key organisations working with multiple global challenges. The organisations working on global challenges vary widely in: 1. What they count as a global challenge 2. How systematically they identify global challenges; and 3. Their emphasis on the most important global challenges For most organisations working with global challenges there are no explanations for the methodology used to select the challenges. Only a few thought leaders, like Tower Watson and their Extreme Risk Report 2013, have a framework for the challenges and estimates of possible impacts. Figure 12: Global challenges that key organisations work with Global Challenges – Twelve risks that threaten human civilisation – The case for a new category of risks 53 2.4 Methodology Climate Change Nuclear War Pandemic Resource Depletion Biological Weapons Computer Failure Government Failure Nanotech Chemical Weapons Artificial Intelligence Genetic Engeneering System Failure Atmospheric Aerosols 13 13 12 9 8 7 7 7 6 5 4 4 0 4 8 12 2 6 10 14 In most cases there is neither a definition of the impact, nor a definition of the probability. The report that focuses on global risk which is probably best known is the WEF Global Risk Report. The WEF’s risk work, with many other groups’, is probably best described as belonging to the category of risk perception rather than risk assessment, where experts are asked to estimate risks, but without any clear definition of probability or impact. The more serious organisations, like the WEF, also clearly define what they do as discussing perception of risk, not a scientific assessment of the actual risk. The WEF describes its perception methodology as follows: “This approach can highlight areas that are of most concern to different stakeholders, and potentially galvanise shared efforts to address them.” 85 The question which people are asked to answer is: “What occurrence causes significant negative impact for several countries and industries?” 86 The respondents are then asked to provide a number on two scales from 1-4, one for impact and another for likelihood (within 10 years).87 It is then up to the respondent to define what 1-4 means, so the major value of the report is to track the changes in perception over the years. Such perception approaches are obviously very interesting and, as the WEF states, can influence actual probability as the readers’ decisions will be influenced by how different challenges are perceived. Still, it is important to remember that the report does not provide an assessment of the actual probability (0-100%) or an assessment of the impact (and not the impact on human suffering, as many respondents likely define risk in monetary terms for their own company or country). An overview of WEF reports from the last ten years indicates that the challenges that likely could happen when applying a five year horizon, like the first signs of climate change, governmental failure and traditional pandemic, are identified. On the other hand, challenges which have very big impacts but lower probability, like extreme climate change, nanotechnology, major volcanoes, AI, and asteroids, tend to get less, or no, attention. An important question to explore is whether a focus on the smaller but still serious impacts of global challenges can result in an increased probability of infinite impacts. For example, there are reasons to believe that a focus on incremental adaptation instead of significant mitigation could be a problem for climate change as it could result in high-carbon lock-in.88 Other research indicates that focus on commercially relevant smaller pandemics could result in actions that make a major pandemic more likely. It is argued that this could happen, for example, by encouraging increased trade of goods while investing in equipment that scans for the type of pandemics that are known. Such a system can reduce the probability for known pandemics while at the same time resulting in an increased probability for new and more serious pandemics.89 Figure 13: The top 12 global challenges that key organisations work with 2.4.3 Workshops global risks 2.5 The list of Two workshops were arranged where the selection of challenges was discussed, one with risk experts in Oxford at the Future of Humanity Institute and the other in London with experts from the financial sector. See Appendix 2 for agenda and participants. In both workshops the list of global challenges was discussed to see if any additional challenges should be included, or if there were reasons to exclude some from the list. No challenge was excluded at the workshops, but one was added. Although little research exists yet that is able to verify the potential impacts, the participants agreed to include Global System Collapse as a risk with possible infinite impact. There was agreement that further research is needed to clarify exactly what parts of the economic and political system could collapse and result in a potentially infinite outcome. The conclusion was that enough research exists to include such a collapse on the list. Based on the risks identified in the literature review and in the review of organisations and applying the criteria for potentially infinite impact, these risks were identified: 1. Extreme Climate Change 2. Nuclear War 3. Global Pandemic 4. Ecological Catastrophe 5. Global System Collapse 6. Major Asteroid Impact 7. Supervolcano 8. Synthetic Biology 9. Nanotechnology 10. Artificial Intelligence (AI) 11. Unknown Consequences 12. Future Bad Global Governance This is an initial list. Additional risks will be added as new scientific studies become available, and some will be removed if steps are taken to reduce their probability90 and/or impact so that they no longer meet the criteria. Four categories of global challenges The challenges included in this report belong to four categories. The first, current challenges, includes those where decisions today can result directly in infinite impacts. They are included even if the time between action and impact might be decades, as with climate change. The second category is exogenous challenges, those where decisions do not – currently – influence probability, but can influence impact. The third category is emerging challenges, those where technology and science are not advanced enough to pose a severe threat today, but where the challenges will probably soon be able to have an infinite impact. The technologies included in emerging challenges, including synthetic biology, nanotechnology and artificial intelligence (AI), will be critical to finding solutions to infinite impacts. Including these technologies should not be seen as an attempt to arrest them. If anything, the development of sustainable solutions should be accelerated. But it is equally important to create guidelines and frameworks to avoid their misuse, whether intentional or accidental. The fourth category, future global policy challenges, is of a different kind. It includes challenges related to the consequences of an inferior or destructive global governance system. This is especially important as well-intended actions to reduce global challenges could lead to future global governance systems with destructive impact. The first category, current challenges, includes: 1. Extreme Climate Change 2. Nuclear War 3. Global Pandemic 4. Ecological Catastrophe 5. Global System Collapse The second category, exogenous challenges, covers: 6. Major Asteroid Impact 7. Supervolcano Those in the third category, emerging challenges, are: 8. Synthetic Biology 9. Nanotechnology 10. Artificial Intelligence (AI) 11. Unknown Consequences The fourth category, global policy challenges, is: 12. Future Bad Global Governance not included 2.5.1 Risks Many risks could severely damage humanity but have not been included in this report. They were excluded for one or more of three reasons: 1. Limited impact. Many challenges can have significant local negative effects, without approaching the “2 billion negatively affected” criterion - tsunamis, for example, and chemical pollution. 2. No effective countermeasures. The report focuses on promoting effective interventions and so ignores challenges where nothing useful can be done to prevent or mitigate the impact, as with nearby gamma-ray bursts. 3. Included in other challenges. Many challenges are already covered by others, or have a damage profile so similar that there seemed no need to have a separate category. Population growth, for one, is an underlying driver significant for climate change and eco-system catastrophe, but without direct large-scale impacts. The challenges mentioned in the reviewed literature and organisations which are not included in this report often refer to economic damage such as “fiscal crises” or “unemployment”. While such impacts could have far-reaching consequences they are obviously of another magnitude than those included here. Some of the risks that were suggested and/or which exist in books and reports about global risks were rejected according to the criteria above. They include: 91 1. Astronomical explosion/nearby gamma-ray burst or supernova.92 These seem to be events of extremely low probability and which are unlikely to be survivable. Milder versions of them (where the source is sufficiently far away) may be considered in a subsequent report. ͢ Not included due to: No effective countermeasures 2. False vacuum collapse. If our universe is in a false vacuum and it collapses at any point, the collapse would expand at the speed of light destroying all organised structures in the universe.93 This would not be survivable. ͢ Not included due to: No effective countermeasures 3. Chemical pollution. Increasingly, there is particular concern about three types of chemicals: those that persist in the environment and accumulate in the bodies of wildlife and people, endocrine disruptors that can interfere with hormones, and chemicals that cause cancer or damage DNA. ͢ Not included due to: Limited impact 4. Dangerous physics experiments creating black holes/strangelets including high energy physics. These risks are of low probability94 and have been subsumed under “Uncertain Risks”. ͢ Not included due to: Included in other challenges 5. Destructive solar flares. Though solar flares or coronal mass ejections could cause great economic damage to our technological civilisation,95 they would not lead directly to mass casualties unless the system lacks basic resilience. They have been subsumed in the Global System Collapse category. ͢ Not included due to: Limited impact/included in other challenges 6. Moral collapse of humanity. Humanity may develop along a path that we would currently find **morally repellent**. The consequences of this are **not clear-cut**, and **depend on value judgements that would be contentious and unshared**.96 Some of these risks (such as global totalitarianism or **enduring poverty**) were included in the Governance Disasters category. ͢ Not included due to: included in other challenges 7. Resource depletion/LULCC/ Biodiversity loss. It has often been argued that declining resources will cause increased conflict.97 Nevertheless such conflicts would **not** be **sufficient in themselves to threaten humanity on a large scale, without a “ System Collapse”** or “Governance Disasters”. ͢ Not included due to: included in other challenge

**Extinction outweighs – any risk is a reason to err neg.**

Seth D. **Baum and** Anthony M. **Barrett 18**. Global Catastrophic Risk Institute. 2018. “Global Catastrophes: The Most Extreme Risks.” Risk in Extreme Environments: Preparing, Avoiding, Mitigating, and Managing, edited by Vicki Bier, Routledge, pp. 174–184.

2. What Is GCR And Why Is It Important? Taken **literally**, a global catastrophe can be any event that is in some way catastrophic across the globe. This suggests a rather low threshold for what counts as a global catastrophe. An event causing just one death on each continent (say, from a jet-setting assassin) could rate as a global catastrophe, because surely these deaths would be catastrophic for the deceased and their loved ones. However, in common usage, a global catastrophe would be **catastrophic** for a significant portion of the globe. Minimum thresholds have variously been set around ten thousand to ten million deaths or $10 billion to $10 trillion in damages (Bostrom and Ćirković 2008), or death of one quarter of the human population (Atkinson 1999; Hempsell 2004). Others have emphasized catastrophes that cause **long-term declines in the trajectory of human civilization** (Beckstead 2013), that human civilization **does not recover from** (Maher and Baum 2013), that drastically reduce humanity’s potential for future achievements (Bostrom 2002, using the term “**existential risk**”), or that result in **human extinction** (Matheny 2007; Posner 2004). A common theme across all these treatments of GCR is that **some catastrophes are vastly more important than others**. Carl Sagan was perhaps the first to recognize this, in his commentary on nuclear winter (Sagan 1983). Without nuclear winter, a global nuclear war might kill several hundred million people. This is obviously a major catastrophe, but humanity would presumably carry on. However, with **nuclear winter**, per Sagan, **humanity could go extinct**. The loss would be not just an additional four billion or so deaths, but the loss of **all future generations**. To paraphrase Sagan, the loss would be billions and billions of lives, or even **more**. Sagan estimated **500 trillion lives**, assuming humanity would continue for ten million more years, which he cited as typical for a successful species. Sagan’s 500 trillion number may even be an **underestimate**. The analysis here takes an adventurous turn, hinging on the evolution of the human species and the long-term fate of the universe. On these long time scales, the descendants of contemporary humans may no longer be recognizably “human”. The issue then is whether the descendants are still worth caring about, whatever they are. If they are, then it begs the question of how many of them there will be. Barring major global catastrophe, Earth will remain habitable for about one billion more years 2 until the Sun gets too warm and large. The rest of the Solar System, Milky Way galaxy, universe, and (if it exists) the multiverse will remain habitable for a lot longer than that (Adams and Laughlin 1997), should our descendants gain the capacity to migrate there. An open question in astronomy is whether it is possible for the descendants of humanity to continue living for an **infinite length of time** or instead merely an **astronomically large but finite** length of time (see e.g. Ćirković 2002; Kaku 2005). Either way, the stakes with global catastrophes **could** be **much larger than the loss of 500 trillion lives.** Debates about the infinite vs. the merely astronomical are of theoretical interest (Ng 1991; Bossert et al. 2007), but they have **limited practical significance**. This can be seen when **evaluating GCRs from a standard risk-equals-probability-times-magnitude framework**. Using Sagan’s 500 trillion lives estimate, it follows that reducing the probability of global catastrophe by a mere one-in-500-trillion chance is of the same significance as saving one human life. Phrased differently, society should **try 500 trillion times harder to prevent a global catastrophe than it should to save a person’s life**. Or, preventing one million deaths is equivalent to a one-in500-million reduction in the probability of global catastrophe. This suggests society should **make extremely large investment in GCR reduction, at the expense of virtually all other objectives.** Judge and legal scholar Richard Posner made a similar point in monetary terms (Posner 2004). Posner used $50,000 as the value of a statistical human life (VSL) and 12 billion humans as the total loss of life (double the 2004 world population); he describes both figures as significant underestimates. Multiplying them gives $600 trillion as an underestimate of the value of preventing global catastrophe. For comparison, the United States government typically uses a VSL of around one to ten million dollars (Robinson 2007). Multiplying a $10 million VSL with 500 trillion lives gives $5x1021 as the value of preventing global catastrophe. But even using “just" $600 trillion, society should be willing to spend at least that much to prevent a global catastrophe, which converts to being willing to spend at least $1 million for a one-in-500-million reduction in the probability of global catastrophe. Thus while reasonable disagreement exists on how large of a VSL to use and how much to count future generations, even low-end positions suggest **vast resource allocations** should be redirected to reducing GCR. This conclusion is only **strengthened** when considering the **astronomical size of the stakes**, but the same point holds either way. The bottom line is that, as long as something along the lines of the standard riskequals-probability-times-magnitude framework is being used, then **even tiny GCR reductions** merit significant effort. This point holds especially strongly for risks of catastrophes that would cause **permanent harm to global human civilization**. The discussion thus far has assumed that all human lives are valued equally. This assumption is **not universally held**. People often value some people more than others, favoring themselves, their family and friends, their compatriots, their generation, or others whom they identify with. Great debates rage on across moral philosophy, economics, and other fields about how much people should value others who are distant in space, time, or social relation, as well as the unborn members of future generations. This debate is crucial for all valuations of risk, including GCR. Indeed, if each of us only cares about our immediate selves, then global catastrophes may not be especially important, and we probably have better things to do with our time than worry about them. While everyone has the right to their **own views and feelings**, we find that the strongest arguments are for the **widely held position** that **all human lives should be valued equally**. This position is succinctly stated in the United States Declaration of Independence, updated in the 1848 Declaration of Sentiments: “We hold these truths to be self-evident: that all men and 3 women are created equal”. Philosophers speak of an agent-neutral, objective “view from nowhere” (Nagel 1986) or a “veil of ignorance” (Rawls 1971) in which each person considers what is best for society **irrespective of which member of society they happen to be**. Such a perspective **suggests valuing everyone equally**, regardless of who they are or where or when they live. This in turn suggests a **very high value for reducing GCR**, or a high degree of priority for GCR reduction efforts.

**AT: Util Bad**

**Consequentialism comes first --- public officials can’t adhere to moral precepts because that prioritizes personal attachments and risks bad outcomes**

**Goodin 95** – professor of government at the University of Essex, and professor of philosophy and social and political theory at Australian National University (Robert E., “Utilitarianism as a Public Philosophy,” Cambridge University Press, Print)BC

As, an Account of the peculiar role responsibilities of public officials (and, by extension, of ordinary individuals in their public capacities as citizens) that vice becomes a virtue, though. Those agents, too, have to come from somewhere, bringing with them a whole raft of baggage of personal attachments, commitments, principles and prejudices. In their public capacities, however, we think it only right and proper that they should stow that baggage as best they can. Complete neutrality might be an impossible ideal. That is another matter." But it seems indisputable that that is an ideal which people in their public capacities should strive to realize as best they are able. That is part (indeed, a central part) of what it is to be a public official ,it all. It is the essence of public service as such that public servants should serve the public at large. Public servants must not play favor­ites. Or consider, again, criticisms revolving around the theme that util­itarianism is a coldly calculating doctrine.23 In personal affairs that is an unattractive feature. There, we would like to suppose that certain sorts of actions proceed immediately from the heart, without much reflection much less any real calculation of consequences. Among in­timates it would be extremely hurtful to think of every kind gesture as being contrived to produce some particular effect. The case of public officials is, once again, precisely the opposite. There, **it is the height of irresponsibility to proceed careless of the consequences**. Public officials are, above all else, obliged to take care: not to go off half cocked, not to let their hearts rule their heads. In Hare's telling example, the very worst thing that might be said of the Suez misadventure was not that the British and French did some per­fectly awful things (which is true, too) but that they did so utterly unthinkingly.24 Related to the critique of utilitarianism as a calculating doctrine is the critique of utilitarianism as a consequentialist doctrine. According to utilitarianism, the effects of an action are everything. There are no actions which are, in and of themselves, morally right or wrong, good or bad. The only things that are good or bad are the effects that actions produce.25 That proposition runs counter to certain ethical intuitions which, at least in certain quarters, are rooted deeply. Those who harbor a Ten Commandments view of the nature of morality see a moral code as being essentially a list of "thou shalts" and "thou shall nots" a list of things that are right or wrong in and of themselves, quite regardless of any consequences that might come from doing them.2" That may or may not be a good way to run one's private affairs.[[1]](#footnote-1)Even those who think it is, however, tend to concede that it is no way to run public affairs. It is in the nature of public officials' role respon­sibilities that they are **morally obliged to "dirty their hands**" - **make hard choices,** do things that are wrong (or would ordinarily be wrong, or would be wrong for ordinary private individuals) in the service of some greater public good.[[2]](#footnote-2) It would be simply irresponsible of public officials (in any broadly secular society, at least) to adhere mindlessly to moral precepts read off some sacred list, literally "whatever the consequences."[[3]](#footnote-3) Doing right though the heavens may fall is not (now­adays, anyway) a particularly attractive posture for public officials to adopt.

**Util is good**

**Greene** 20**10** – Joshua, Associate Professor of Social science in the Department of Psychology at Harvard University (The Secret Joke of Kant’s Soul published in Moral Psychology: Historical and Contemporary Readings, accessed: www.fed.cuhk.edu.hk/~lchang/material/Evolutionary/Developmental/Greene-KantSoul.pdf)

What **turn-of-the-millennium science** is telling us is that human **moral judgment is not a pristine rational enterprise**, that our moral judgments are driven by a hodgepodge of emotional dispositions, which themselves were shaped by a hodgepodge of evolutionary forces, both biological and cultural. Because of this, it is **exceedingly unlikely that there is any rationally coherent normative moral theory that can accommodate our moral intuitions**. Moreover, anyone who claims to have such a theory, or even part of one, **almost certainly doesn't**. Instead, what that person probably has is a moral rationalization. It seems then, that we have somehow crossed the infamous "is"-"ought" divide. How did this happen? Didn't Hume (Hume, 1978) and Moore (Moore, 1966) warn us against trying to derive an "ought" from and "is?" How did we go from descriptive scientific theories concerning moral psychology to skepticism about a whole class of normative moral theories? The answer is that we did not, as Hume and Moore anticipated, attempt to derive an "ought" from and "is." That is, our method has been inductive rather than deductive. We have inferred on the basis of the available evidence that the phenomenon of rationalist deontological philosophy is best explained as a rationalization of evolved emotional intuition (Harman, 1977). Missing the Deontological Point I suspect that rationalist deontologists will remain unmoved by the arguments presented here. Instead, I suspect, they will insist that I have **simply misunderstood what** Kant and like-minded **deontologists are all about**. Deontology, they will say, isn't about this intuition or that intuition. It's not defined by its normative differences with consequentialism. Rather, deontology is about taking humanity seriously. Above all else, it's about respect for persons. It's about treating others as fellow rational creatures rather than as mere objects, about acting for reasons rational beings can share. And so on (Korsgaard, 1996a; Korsgaard, 1996b). This is, no doubt, how many deontologists see deontology. But this insider's view, as I've suggested, **may be misleading**. The problem, more specifically, **is that it defines deontology in terms of values that are not distinctively deontological**, though they may appear to be from the inside. Consider the following analogy with religion. When one asks a religious person to explain the essence of his religion, one often gets an answer like this: "It's about love, really. It's about looking out for other people, looking beyond oneself. It's about community, being part of something larger than oneself." This sort of answer accurately captures the phenomenology of many people's religion, but it's nevertheless inadequate for distinguishing religion from other things. This is because many, if not most, non-religious people aspire to love deeply, look out for other people, avoid self-absorption, have a sense of a community, and be connected to things larger than themselves. In other words, secular humanists and atheists can assent to most of what many religious people think religion is all about. From a secular humanist's point of view, in contrast, what's distinctive about religion is its commitment to the existence of supernatural entities as well as formal religious institutions and doctrines. And they're right. These things really do distinguish religious from non-religious practices, though they may appear to be secondary to many people operating from within a religious point of view. In the same way, I believe that most of the standard deontological/Kantian self-characterizatons **fail to distinguish deontology from other approaches to ethics**. (See also Kagan (Kagan, 1997, pp. 70-78.) on the difficulty of defining deontology.) It seems to me that consequentialists, as much as anyone else, **have respect for persons**, are **against treating people as mere objects,** wish **to act for reasons that rational creatures can share**, etc. A consequentialist respects other persons, and refrains from treating them as mere objects, by **counting every person's well-being in the decision-making process**. Likewise, a consequentialist attempts to act according to reasons that rational creatures can share by acting according to principles that **give equal weight to everyone's interests**, i.e. that are impartial. This is not to say that consequentialists and deontologists don't differ. They do. It's just that the real differences may not be what deontologists often take them to be. What, then, distinguishes deontology from other kinds of moral thought? A good strategy for answering this question is to start with concrete disagreements between deontologists and others (such as consequentialists) and then work backward in search of deeper principles. This is what I've attempted to do with the trolley and footbridge cases, and other instances in which deontologists and consequentialists disagree. If you ask a deontologically-minded person why it's wrong to push someone in front of speeding trolley in order to save five others, you will get characteristically deontological answers. Some **will be tautological**: **"Because it's murder!"** Others will be more sophisticated: "The ends don't justify the means." "You have to respect people's rights." **But**, as we know, **these answers don't really explain anything**, because if you give the same people (on different occasions) the trolley case or the loop case (See above), **they'll make the opposite judgment**, even though their initial explanation concerning the footbridge case applies equally well to one or both of these cases. Talk about rights, respect for persons, and reasons we can share are natural attempts to explain, in "cognitive" terms, what we feel when we find ourselves having emotionally driven intuitions that are odds with the cold calculus of consequentialism. Although these explanations are inevitably incomplete, **there seems to be "something deeply right" about them because they give voice to powerful moral emotions**. But, as with many religious people's accounts of what's essential to religion, they don't really explain what's distinctive about the philosophy in question.

**Prefer utilitarian approaches to policy that cast aside empathy --- valuing how discussions make us feel rather than the truth of the discussion ensures disastrous decision-making**

**Bloom 14** [Paul Bloom, Brooks and Suzanne Ragen Professor of Psychology & Cognitive Science at Yale University, “Against Empathy, Boston Review, September 13, 2014, <http://bostonreview.net/forum/paul-bloom-against-empathy>]

When asked what I am working on, I often say I am writing a book about empathy. People tend to smile and nod, and then I add, “**I’m against it**.” This usually gets an uncomfortable laugh.

This reaction surprised me at first, but I’ve come to realize that taking a position against empathy is like announcing that you hate kittens—a statement so outlandish it can only be a joke. And so I’ve learned to clarify, to explain that I am not against morality, compassion, kindness, love, being a good neighbor, doing the right thing, and making the world a better place. My claim is actually the opposite: if you want to be good and do good, empathy is a poor guide.

The word “empathy” is used in many ways, but here I am adopting its most common meaning, which corresponds to what eighteenth-century philosophers such as Adam Smith called “sympathy.” It refers to the process of experiencing the world as others do, or at least as you think they do. To empathize with someone is to put yourself in her shoes, to feel her pain. Some researchers also use the term to encompass the more coldblooded process of assessing what other people are thinking, their motivations, their plans, what they believe. This is sometimes called “cognitive,” as opposed to “emotional,” empathy. I will follow this convention here, but we should keep in mind that the two are distinct—they emerge from different brain processes; you can have a lot of one and a little of the other—and that most of the discussion of the moral implications of empathy focuses on its emotional side.

Some degree of emotional empathy is bred in the bone. The sight and sound of another’s suffering is unpleasant for babies and, as soon as they are mobile enough, they try to help, patting and soothing others in distress. This is not uniquely human: the primatologist Frans de Waal notes that chimps will often put their arms around the victim of an attack and pat her or groom her.

Empathy can occur automatically, even involuntarily. Smith describes how “persons of delicate fibres” who notice a beggar’s sores and ulcers “are apt to feel an itching or uneasy sensation in the correspondent part of their own bodies.” John Updike writes, “My grandmother would have choking fits at the kitchen table, and my own throat would feel narrow in sympathy.”

And empathy can be extended through the imagination. In a speech before he became president, Barack Obama stressed how important it is

to see the world through the eyes of those who are different from us—the child who’s hungry, the steelworker who’s been laid off, the family who lost the entire life they built together when the storm came to town. . . . When you think like this—when you choose to broaden your ambit of concern and empathize with the plight of others, whether they are close friends or distant strangers—it becomes harder not to act, harder not to help.

Obama is right about this last part; there is considerable support for what the psychologist C. Daniel Batson calls “the empathy-altruism hypothesis”: when you empathize with others, you are more likely to help them. In general, empathy serves to dissolve the boundaries between one person and another; it is a force against selfishness and indifference.

It is easy to see, then, how empathy can be a moral good, and it has many champions. Obama talks frequently about empathy; witness his recent claim, after his first meeting with Pope Francis, that “it’s the lack of empathy that makes it very easy for us to plunge into wars. It’s the lack of empathy that allows us to ignore the homeless on the streets.” In The Empathetic Civilization (2009) Jeremy Rifkin argues that the only way our species will survive war, environmental degradation, and economic collapse is through the enhancement of “global empathy.” This past June, Bill and Melinda Gates concluded their Stanford commencement address by asking students to nurture and expand their empathetic powers, essential for a better world.

Most people see the benefits of empathy as akin to the evils of racism: too obvious to require justification. I think this is a mistake. I have argued elsewhere that certain features of empathy make it a poor guide to social policy. **Empathy is biased**; we are more prone to feel empathy for attractive people and for those who look like us or share our ethnic or national background. And empathy is narrow; it connects us to particular individuals, real or imagined, but is **insensitive to numerical differences and statistical data**. As Mother Teresa put it, “If I look at the mass I will never act. If I look at the one, I will.” Laboratory studies find that we really do care more about the one than about the mass, so long as we have personal information about the one.

In light of these features, **our public decisions will be fairer and more moral once we put empathy aside.** Our **policies are improved when we appreciate that a hundred deaths are worse than one**, even if we know the name of the one, and when we acknowledge that the life of someone in a faraway country is worth as much as the life a neighbor, even if our emotions pull us in a different direction. **Without empathy**, **we are better able to grasp the importance of vaccinating children and responding to climate change**. These acts impose costs on real people in the here and now for the sake of **abstract future benefits**, so tackling them **may require overriding empathetic responses** that favor the comfort and well-being of individuals today. We can rethink humanitarian aid and the criminal justice system, choosing to draw on a reasoned, even **counter-empathetic**, **analysis of moral obligation and likely consequences**.

**The presence of intervening actors doesn’t decrease the moral responsibility for your own actions**

**Cummiskey, 99** (David, professor of philosophy at Bates, Gewirth, ed: Boylan, p. 134-135)

Indeed, the Principle of the Intervening Action simply asserts what must be shown. As Gewirth emphasizes, all agents have a right not to be killed. The question at issue is whether it is sometimes obligatory to kill to prevent more killings. If one responds that it is not because another agent is doing the killing, then one is simply assuming that the duty in question is an agent-relative restriction. This unargued assertion, however, conflicts with the basic PGC requirement to respect the rights of all persons affected by one’s choice. Since the persons being killed have a right not to be killed, and since they are going to be killed as a result of my refusal, they are indeed recipients of my action in the sense that is decisive as far as morality is concerned. I thus must act in light of the equal rights of all who will be killed if I refuse and of the person who will be killed if I give in to the terrorists. As we have seen above, all other things equal, the rights of all will be more secure if we accept a principle where the rights of a few give way to the rights of many. The consequentialist criterion of the degree of needfulness for action requires a cost-benefit analysis of the options. Since the fewer people killed the more objective basic needs are protected, it appears that the right of the many not to be killed outweigh the right of the one not to be killed. Much more could be said about the principle of intervening action. For present purposes, I only wish to focus our attention on the central point: Gewirth has not provided an adequate argument for this principle and, in the final analysis, the “criterion of degree of needfulness for action,” which follows directly from the PGC, must determine which course of action is unconditionally obligatory. And this consequentialist criterion, in principle, could clearly justify killing some to prevent others from being killed. So the PGC does not provide a basis for Gewirth’s alleged absolute right not to be killed as a means of preventing a nuclear catastrophe. Indeed, in principle, the PGC could even justify killing one as a means of preventing two others from being killed—if all other considerations are indeed equal.

**Predictions**

**Predictions of high-magnitude impacts are accurate and valuable.**

**Gleditsch 12** (Kristian S. Gleditsch, Department of Government, University of Essex & Peace Research Institute Oslo, and Michael D. Ward, Department of Political Science, Duke, 2012, “Forecasting is difficult, especially about the future: Using contentious issues to forecast interstate disputes,” Journal of Peace Research, 50(1) 17–31, Sage Journals)  
There have been **remarkably few efforts** to generate global forecasts or risk profiles for interstate conflict. Moreover, the most prominent efforts to consider the predictive ability of models of interstate conflicts have based their research on models that were not actually proposed with forecasting in mind. A notable example here is Beck, King & Zeng (2000), who essentially adopt the so-called liberal peace model of Russett & Oneal (2001). Certainly, nothing akin to the Political Instability Task Force’s now annual projections (beginning with Gurr, Marshall & Khosla, 2000) exists for international conflicts. Perhaps not surprisingly, many observers are **very skeptical** of the ability of academic researchers to anticipate conflict between states, at least beyond very short time horizons. Research in recent decades has seen a large number of hypotheses generated to explain under what conditions militarized interstate conflict is more or less likely. This avenue of research has been primarily inspired by research on the so-called democratic peace, or the absence of conflict between democracies. Indeed, there are thousands of scholarly works mentioning the term militarized interstate dispute (MID), most of which use these data for some kind of empirical examination of a proposition about disputes. Yet, the evidence suggests that the ability of this body of work to forecast conflict out-of-sample is **decidedly disappointing**. Ward, Siverson & Cao (2007) found that most of the recent statistical studies of militarized interstate disputes in prominent political science and international relations journals were unable to predict the outbreak of a single dispute out-of-sample (see also Beck, King & Zeng, 2000). Many researchers have sought to improve on the ability to forecast militarized interstate conflict by turning to alternative statistical methods. Beck, King & Zeng (2000), for example, find that neural networks perform marginally better than generalized linear regression models in forecasting conflict from the same input factors.1 Changes in estimation methods or statistical techniques per se, however, have at best led only to limited improvements in out-of-sample predictive ability. Our argument is that **simply identifying inappropriate methods** as the key source of the problem in forecasting conflict may give us the **wrong diagnosis** and lead us down less productive avenues. A more fundamental problem is models that provide a poor basis for forecasting by disregarding the motives for conflict to arise, or by only considering motives in a relatively superficial manner. Models that have been proposed for research on the democratic peace, notably the work of Russett & Oneal (2001), are primarily intended to examine whether certain characteristics of liberal institutions, such as democracy and trade, make conflict on average less likely relative to baseline risks of conflict. Although these approaches may be appropriate for testing the original propositions of interest, they essentially ignore the **contentious issues** that might cause states to resort to violence and instead treat these contentious issues as exogenous features, typically hidden inside a so-called ‘black box’ of the baseline risk of conflict. Our own initial foray into out-of-sample prediction for a state-level model indicates that spatial information about other conflict events can help to improve forecasts (see Ward & Gleditsch, 2002). Although this allows predictions of conflict to be conditional on other observed events rather than treating each conflict as an independent observation, the approach still ignores the issues over which such conflicts may have arisen initially. We believe that greater attention to the **specific reasons for the occurrence of conflicts** and the incompatibilities that may generate the use of violence can help **improve our ability to forecast conflict.** Although we recognize that different models may be appropriate to evaluate particular propositions and to forecast events, in our view **the enterprise of prediction has great potential** for winnowing bad ideas out of theories on the causes of conflict and avoiding the problem of retrospective biases in conventional hypothesis testing on the data used to develop the hypotheses in the first place (see Ward, Greenhill & Bakke, 2010). In fairness, much of the existing work on the statistical modeling of conflict has bypassed motivation since it is genuinely difficult to establish what states fight over and what their possible motivation for fighting might be. Nevertheless, **the fact that something is difficult to evaluate does not mean that simply ignoring it is the best course of action**. Another tradition in research on conflict has sought to identify incompatibilities in terms of contentious issues, such as territorial or maritime claims (Diehl, 1992; Mansbach & Vasquez, 1981). Recent efforts to examine these propositions empirically have found considerable evidence that cases where such claims exist are more likely to see militarized activities (Hensel, 2001; Hensel et al., 2008; Hensel & Mitchell, 2010). Even so, at present this line of research has primarily engaged in testing hypotheses about whether coefficient estimates are significantly different from 0 or the in-sample post-diction of conflicts, and has not yet examined if information on contentious issues may be helpful for forecasting dyadic conflict out-of-sample. Here we explicitly consider whether taking into account information on contentious issues and conflict management can help improve on forecasting interstate conflict and our understanding of conflict dynamics. Although we focus on statistical approaches to interstate conflict in this article, many of our arguments also apply to problems in traditional theories of conflict and qualitative approaches to prediction or anticipating political events (see Tetlock, 2005). Traditional theories of interstate conflict tend to focus on **structural features** presumed to influence the opportunities for conflict such as the distribution of power in the international system or relative balance of power (see e.g. Waltz, 1979). These theories display little interest in the specific incompatibilities that may motivate the use of violence. However, structural factors rarely change rapidly, but violent conflict tends to be episodic, and hence cannot be adequately explained merely by reference to permissive conditions (see Fearon, 1995). Likewise, our core argument applies to studies of civil war, which tend to emphasize opportunities for conflict rather than motivations for conflict (see Cederman, Weidmann & Gleditsch, 2011), and where evidence for the predictive ability of existing statistical efforts seems similarly disappointing (see Ward, Greenhill & Bakke, 2010). Many political and area study experts, typically using informal methods for deriving predictions, often have strong confidence in their ability to forecast events. However, the comprehensive series of studies by Tetlock (2005), who asked experts to rate a series of outcomes which could then be compared against the historical record, provide little support for the forecasting ability of political experts.

**Predictions are accurate and good**

Michael D. **Ward 13**, Professor of Political Science at Duke University, Niles W. Metternich, University of College London, Cassy L. Dorff, Max Gallop, Florian M. Hollenbach, Anna Schultz, and Simon Weschle, "Learning from the Past and Stepping into the Future: Toward a New Generation of Conflict Prediction", International Studies Review (2013) 15, 473-490

**Political events are frequently framed as unpredictable. Who could have predicted the Arab Spring, 9/11, or the end of the cold war?** This skepticism about prediction reflects an underlying desire to forecast. **Predicting political events is difficult because they result from complex social processes**. **However, in recent years, our capacity to collect information** **on social behavior and our ability to process large data have increased to degrees only foreseen in science fiction. This new ability to analyze and predict behavior confronts a demand for better political forecasts that may serve to inform and even help to structure effective policies in a world in which prediction in everyday life has become commonplace. Only a decade ago, scholars interested in civil wars undertook their research with constrained resources**, limited data, and statistical estimation capabilities that seem underdeveloped by current standards. **Still, major advances did result** from these efforts. **Consider “Ethnicity, Insurgency and Civil War” by Fearon** and Laitin (2003), **one of the most venerated and cited articles** about the onset of civil wars. Published in 2003, it has over 3,000 citations in scholar.google.com and almost 900 citations in the Web of Science (as of April 2013). It has been cited prominently in virtually every social science discipline in journals ranging from Acta Sociologica to World Politics; and it is the most downloaded article from the American Political Science Review.2 ¶ **This article is rightly regarded as an important, foundational piece of scholarship. However, in** the summer of **2012, it was used by** Jacqueline **Stevens** in a New York Times Op-Ed **as evidence that political scientists are bad forecasters**. **That claim was wildly off the mark in that Fearon and Laitin do not focus on forecasting, and Stevens ignored other, actual forecasting efforts in political science**. Stevens’ point—which was taken up by the US Congress—was that government funding on quantitative approaches was being wasted on efforts that did not provide accurate policy advice. **In contrast to Stevens, we argue that conflict research in political science can be substantially improved by more**, not less, **attention to predictions through quantitative approaches**.¶ We argue that the **increasing availability of disaggregated data and advanced estimation techniques are making forecasts of conflict more accurate and precise, thereby helping to evaluate the utility of different models and winnow the good from the bad.** **Forecasting also helps to prevent overfitting and reduces confirmation bias**. As such, **forecasting efforts can be used to help validate models, to gain greater confidence in the resulting estimates, and to ultimately present robust models that may allow us to improve the interaction with decision makers seeking greater clarity about** the **implications of potential actions**.

**Turn—rejecting predictions makes them inevitable—decisionmakers will rely on ideology instead of analysis**

**Fitzsimmons 7** (Michael, “The Problem of Uncertainty in Strategic Planning”, Survival, Winter 06/07)

But handling even this weaker form of uncertainty is still quite challeng- ing. If not sufficiently bounded, a high degree of variability in planning factors can exact a significant price on planning. The complexity presented by great variability strains the cognitive abilities of even the most sophisticated decision- makers.15 And even a robust decision-making process sensitive to cognitive limitations necessarily sacrifices depth of analysis for breadth as variability and complexity grows. It should follow, then, that in planning under conditions of risk, variability in strategic calculation should be carefully tailored to available analytic and decision processes. Why is this important? What harm can an imbalance between complexity and cognitive or analytic capacity in strategic planning bring? Stated simply, where analysis is silent or inadequate, **the personal beliefs of decision-makers** **fill the void**. As political scientist Richard Betts found in a study of strategic sur- prise, in ‘an environment that lacks clarity, abounds with conflicting data, and allows no time for rigorous assessment of sources and validity, ambiguity allows intuition or wishfulness to drive interpretation ... The greater the ambiguity, the greater the impact of preconceptions.’16 The decision-making environment that Betts describes here is one of political-military crisis, not long-term strategic planning. But a strategist who sees uncertainty as the central fact of his environ- ment brings upon himself some of the pathologies of crisis decision-making. He invites ambiguity, takes conflicting data for granted and **substitutes a priori scepticism about the validity of prediction** for time pressure as a rationale for discounting the importance of analytic rigour. It is important not to exaggerate the extent to which data and ‘rigorous assessment’ can illuminate strategic choices. Ambiguity is a fact of life, and scepticism of analysis is necessary. Accordingly, the intuition and judgement of decision-makers will always be vital to strategy, and attempting to subordinate those factors to some formulaic, deterministic decision-making model would be both undesirable and unrealistic. All the same, there is danger in the opposite extreme as well. Without careful analysis of what is relatively likely and what is relatively unlikely, what will be the possible bases for strategic choices? A decision-maker with no faith in prediction is left with little more than a set of worst-case scenarios and his existing beliefs about the world to confront the choices before him. Those beliefs may be more or less well founded, but if they are not made explicit and subject to analysis and debate regarding their application to particular strategic contexts, they remain only beliefs and premises, rather than rational judgements. Even at their best, such decisions are likely to be poorly understood by the organisations charged with their implementation. At their worst, such decisions may be poorly understood by the decision-makers themselves.

**AT: Bernstein 2k**

**Bernstein is straw person logic—impact scenarios aren’t insular ‘*point predictions’*—they’re illustrations of plausible IR outcomes—at worst, their K means you just prefer specificity**

**Chernoff 5**—IR prof, Colgate. PhD from Yale. (Fred, The Power of International Theory, Ed. Barry Buzan and Richard Little, 157-9, AMiles)

In the social sciences the explanations of state behaviour in rational-choice theory, which posits the existence of a rationally behaving state, or in economics, which posits the existence of ‘the rational economic person’, are clearly idealisations. However, these are parallel to – and not completely disconnected from – the postulation of laws governing frictionless machines or ideal gases in the physical sciences. With regard to underlying and immediate causes, Bernstein et al. advance a number of points in the course of offering the critique of ‘predictiveness’ of IR theory and the scenario-based alternative that Bernstein et al. present. The main point to be made with respect to the use made by Bernstein et al. of the ‘uncertain relationships between underlying and immediate causes’ is that, in their view, such a connection ‘makes point prediction extraordinarily difficult’ (2000: 47). As will become clear, this chapter argues that IR theorists offer non-point predictions and that such predictions are often of genuine policy value. A statement like ‘a rapid withdrawal of US military forces from Iraq after the defeat of Saddam Hussein is likely to lead to increased violence and long-term instability’, would not qualify as a point prediction. Yet it is the sort of prediction that policymakers typically rely on and it has value for policy-makers. Bernstein et al. thus commit a straw-**person fallacy**, as do other critics of ‘predictiveness’ discussed above. They raise the bar of ‘prediction’ much too high. The sort of prediction that policy-makers need is often much less than ‘point prediction’. One may grant Bernstein et al. all of their points seeking to undermine point prediction and still conclude that they have not provided good reason to reject many sorts of statements that satisfy the requirements of prediction, as laid out in the definition on p. 8. With regard to learning, IR differs from predictive natural sciences in that, according to Bernstein et al., ‘[m]olecules do not learn from experience. People do, or think they do. … We know that expectations and behavior are influenced by experience, one’s own and others’ (2000: 47). They argue that US policies visàvis the USSR were a response to the failure of appeasement in the 1930s, and those policies were a response to British leaders’ belief that more aggressive policies failed to keep peace in 1914. They cite concepts like ‘chain reactions’ and ‘contagion effects’ to describe these phenomena and hazard analysis for their measurement. But they charge that these do not succeed in explaining ‘how and why these patterns emerge and persist’ (Bernstein et al. 2000: 47). They note also that theories that attempt to predict the future predict incorrectly in part because groups (often states) react in such a way as to prevent their predictions from obtaining. ‘Human prophecies … are often self-negating’ (Bernstein et al. 2000: 52). Another approach that has considerable explanatory success is cybernetic theory, which lays out mechanisms and may even be viewed as predictive. Cybernetic theory was developed by Wiener (1949) and applied to IR especially by Deutsch et. al (1957), and specifically to foreign-policy decision-making by Steinbruner (1974), Chernoff (1995) and others. It not only accords with the data but offers a causal mechanism to account for the patterns. Bernstein et al. argue that actors can change ‘the rules of the game’ and consequently ‘general theories of process in international relations will have restricted validity’ (Bernstein et al. 2000: 52). Generalisations will apply only to ‘discrete portions’ of history. They add that ‘scholars need to specify carefully the temporal and geographic domains to which their theories are applicable. We suspect those domains are often narrower and more constrained than is generally accepted’ (Bernstein et al. 2000: 52). This is, however, just what one of the most systematic and generalising of all IR scholars, i.e., Waltz, demands. He maintains that a body of propositions does not even qualify as a theory unless it so specifies. Once these restrictions on domain are specified, theoretical generalisations might have value. For example, Waltz says that whether we are interested in natural sciences or social sciences, ‘[n]o matter what the subject, we have to bound the domain of our concern, to organize it, to simplify the materials we deal with, to concentrate on the central tendencies, and to single out the strongest propelling forces’ (1979: 68). He later adds, ‘[t]o be a success … a theory has to show how international politics can be conceived of as a domain distinct from economic, social, and other international domains that one may conceive of ’ (1979: 79). So it is not accurate to suggest that IR theorists, at least careful ones, do not bound or constrain the scope of their studies and generalisations. While bounded theories may offer predictions, the fallible nature of all empirical knowledge and the probabilistic nature of IR laws does restrict what can be predicted. The longer the chain of reasoning and the greater the number of probabilistic propositions that are conjoined, the less one may rely on the predicted event. But this is an argument for limitations on certain types of predictions, not an argument against the predictiveness of IR or social science theory. Bernstein et al. note also that there is the problem of the ‘single case’. They correctly observe that policy-makers often worry about a single instance of an event type (e.g., possible war with our neighbours on our western frontier in the next year) rather than with general propositions (e.g., the problem of war in general, or great-power war, etc.). First, with regard to the relevance of theory to policy, it is important to note that policy-makers do sometimes care about longrun patterns or generalisations in some of their decision-situations. Long-run generalisations are very frequently important for policy-makers. For example, in the 1990s, numerous new states emerged in Europe and central Asia due to the collapse of the Soviet Union and the end of its domination of central eastern Europe. Policy-makers had to understand as correctly as possible the truth of generalisations in DP theory in order to decide how much in the way of financial and diplomatic resources should be committed to help promote democratic polities in the region. Policy-makers in 1994 were not primarily worried about a specific war, say, between Hungary and Slovakia, but rather the long-run chances of conflict arising in a Europe with many nondemocratic states as compared to a Europe with very few such states. Policy-makers could consider two alternative scenarios of the future. One is a Europe with some democratic states and many non-democratic ones, possibly including a non-democratic Russia. The other one envisions a Europe with virtually all democratic states. The generalisations about the effects of democracy on behaviour would allow policy-makers to draw general conclusions about peace and international stability from each scenario.

**AT: Structural Violence**

**War turns structural violence and no tradeoff**

John **Horgan 14**. Director of the Center for Science Writings at Stevens Institute of Technology, “To End War, Focus on Culture Rather than "Root Causes"”, Scientific American, 8-18, https://blogs.scientificamerican.com/cross-check/to-end-war-focus-on-culture-rather-than-root-causes/

When I started researching war, I also assumed that to get rid of war, we have to get rid of its root causes. The trouble is, scholars have identified countless causes of war. One pseudo-explanation (which I'm glad Kloor does not mention, and which I rebut early on in my book and in posts such as this) is that war stems from a compulsion bred into our ancestors by natural selection. Biology underpins war, as it underpins all human behaviors. The crucial question is, why does war break out in certain places and times and not others? The most popular non-biological explanations of war are what I call the **Malthusian** and **Marxist** hypotheses. The first posits that war stems from our tendency to over-reproduce and hence fight over land and other resources. The second holds that war stems from inequality, the tendency of societies (especially capitalist ones) to divide into haves and have-nots. Scholars have also blamed wars on **religion**, **racism** and **nationalism**, which Kloor mentions above, as well as such **fundamental social traits** as hierarchy, sexism and injustice. If you **cherry pick**, you **always find evidence** to support your favorite theory. But as scholars such as Lewis Fry Richardson (whom my friend David Berreby recently profiled) have shown, neither the Malthusian and Marxist theories **no**r any of the other **explanations above can account for the vast diversity of wars**. Moreover, some factors that provoke conflict, such as religion, can also inhibit it. Religion has inspired some of our greatest antiwar leaders, notably Gandhi and Martin Luther King. I have found only one theory of war that fits the facts. The theory holds that war is a self-perpetuating, contagious meme, which can propagate **independently of other social and environmental factors**. As anthropologist Margaret Mead put it in a famous 1940 essay, "Warfare Is Only an Invention—Not a Biological Necessity." In other words, the major cause of war is war itself, which has a terrible tendency to spread even to societies that would prefer to remain peaceful. I make this point in my book and in a 2010 blog post, "Margaret Mead’s war theory kicks butt of neo-Darwinian and Malthusian models." Here is an edited excerpt: In his 1997 book War Before Civilization, anthropologist Lawrence Keeley notes that war among North American Indians often stemmed from the aggression of just a few extremely warlike tribes, "rotten apples that spoiled their regional barrels." He added, "Less aggressive societies, stimulated by more warlike groups in their vicinity, become more bellicose themselves." Societies in a violent region, the political scientist Azar Gat emphasizes in his 2006 book War in Human Civilization, have a strong incentive to carry out preemptive attacks. Societies may "attack the other side in order to eliminate or severely weaken them as a potential enemy. Indeed, this option only makes the other side more insecure, rendering the security dilemma more acute. War can thus become a self-fulfilling prophecy. The fear of war breeds war." Many people are **pessimistic** about ending war because they assume it will require **radical social engineering**. World peace will require eliminating poverty, inequality, sexism, **racism** or **[fill in the blank]**. We will need to eradicate religion, or all embrace the same religion. We will need to get rid of all nation states and become anarchists, or form a single global government. My analysis of war suggests that if we want to end war, **we don't need to create a society radically different** from our own, let alone a utopia. If we want to end war, we should focus on ending war and the culture of war rather than on **supposed causal factors**. If we can do that, we will **take a major step toward solving many** of our **other social problems**, as I argued in my previous post. And that brings me to Keith Kloor's final challenge to me. He devotes much of his column to a discussion of how extremists on both sides of the conflict between Israel and Palestine have "hijacked the peace process. Horrific spasmodic cycles of violence and death is the result." He asks me how we can "rid the world of extremist groups that sow the seeds of war." Kloor has his **causation backwards**. Just as war promotes poverty, tyranny, inequality and resource depletion **at least as much as vice versa**, so war promotes fanaticism. Once militarism seizes hold of a society, it can transform vast populations into virtual sociopaths. It turns decent, ethical, reasonable people into intolerant fanatics capable of the most heinous acts. Breaking out of what Kloor calls "spasmodic cycles of violence and death" can be extraordinarily difficult, but history offers many examples of societies that have done just that. Germany and France were bitter, bloody rivals for centuries. But it is now inconceivable that Germany and France—or any members of the European Union--would go to war against each other. One of my favorite examples of a nation that has renounced militarism is Costa Rica. Like many of its neighbors in Central American, Costa was once wracked by terrible violence. But after a bloody civil war in the 1940s, Costa Rica disbanded its army, freeing up more funds for education, health care, transportation and tourism. It is often ranked as one of the most peaceful, healthy, "happy" nations in the world.

**Root cause explanations of international politics don’t exist. Their attempt to impose a universal root cause results in unchecked reification and a larger example of what they’re critiquing.**

**Bleiker 14** – (6/17, Roland, Professor of International Relations at the University of Queensland, “International Theory Between Reification and Self-Reflective Critique,” International Studies Review, Volume 16, Issue 2, pages 325–327)

This book is part of an increasing trend of scholarly works that have embraced poststructural critique but want to ground it in more positive political foundations, while retaining a reluctance to return to the positivist tendencies that implicitly underpin much of constructivist research. The path that Daniel Levine has carved out is innovative, sophisticated, and convincing. A superb scholarly achievement.

For Levine, the key challenge in international relations (**IR**) scholarship is what he calls **“unchecked reification”**: the widespread and dangerous process of forgetting “the distinction between theoretical concepts and the real-world things they mean to describe or to which they refer” (p. 15). The dangers are real, Levine stresses, because **IR deals with** some of the most difficult issues, from **genocides to war**. Upholding one subjective position without critical scrutiny can thus have **far-reaching consequences**. Following Theodor Adorno—who is the key theoretical influence on this book—Levine takes a post-positive position and assumes that the world cannot be known outside of our human perceptions and the values that are inevitably intertwined with them. His ultimate goal is to overcome reification, or, to be more precise, to recognize it as an inevitable aspect of thought so that its dangerous consequences can be mitigated.

Levine proceeds in three stages: First he reviews several decades of IR theories to resurrect critical moments when scholars displayed an acute awareness of the dangers of reification. He refreshingly breaks down distinctions between conventional and progressive scholarship, for he detects self-reflective and critical moments in scholars that are usually associated with straightforward positivist positions (such as E.H. Carr, Hans Morgenthau, or Graham Allison). But Levine also shows how these moments of self-reflexivity never lasted long and were driven out by the compulsion to offer systematic and scientific knowledge.

The second stage of Levine's inquiry outlines why IR scholars regularly closed down critique. Here, he points to a range of factors and phenomena, from peer review processes to the speed at which academics are meant to publish. And here too, he eschews conventional wisdom, showing that work conducted in the wake of the third debate, while explicitly post-positivist and critiquing the reifying tendencies of existing IR scholarship, often lacked critical self-awareness. As a result, Levine believes that many of the respective authors failed to appreciate sufficiently that “**reification is a consequence of all thinking**—including itself” (p. 68).

The third objective of Levine's book is also the most interesting one. Here, he outlines the path toward what he calls “sustainable critique”: a form of self-reflection that can counter the dangers of reification. Critique, for him, is not just something that is directed outwards, against particular theories or theorists. It is also inward-oriented, ongoing, and sensitive to the “limitations of thought itself” (p. 12).

The challenges that such a sustainable critique faces are formidable. Two stand out: First, if the natural tendency to forget the origins and values of our concepts are as strong as Levine and other Adorno-inspired theorists believe they are, then how can we actually recognize our own reifying tendencies? Are we not all inevitably and subconsciously caught in a web of meanings from which we cannot escape? Second, if one constantly questions one's own perspective, does one not fall into a relativism that loses the ability to establish the kind of stable foundations that are necessary for political action? Adorno has, of course, been critiqued as relentlessly negative, even by his second-generation Frankfurt School successors (from Jürgen Habermas to his IR interpreters, such as Andrew Linklater and Ken Booth).

The response that Levine has to these two sets of legitimate criticisms are, in my view, both convincing and useful at a practical level. He starts off with depicting reification not as a flaw that is meant to be expunged, but as an a priori condition for scholarship. The challenge then is not to let it go unchecked.

**Methodological pluralism lies at the heart of Levine's sustainable critique**. He borrows from what Adorno calls a “constellation”: an attempt to juxtapose, rather than integrate, different perspectives. It is in this spirit that Levine advocates **multiple methods to understand the same event or phenomena**. He writes of the need to validate “multiple and mutually incompatible ways of seeing” (p. 63, see also pp. 101–102). In this model, a scholar oscillates back and forth between different methods and paradigms, trying to understand the event in question from multiple perspectives. **No single method can ever adequately represent the event or should gain the upper hand**. But each should, in a way, **recognize and capture details or perspectives that the others cannot** (p. 102). In practical terms, this means **combining a range of methods** even when—**or, rather, precisely when**—they are deemed incompatible. They can **range from poststructual deconstruction** to the tools pioneered and championed by **positivist social sciences**.

The benefit of such a **methodological polyphony** is not just the opportunity to bring out nuances and new perspectives. Once the false hope of a smooth synthesis has been abandoned, the very incompatibility of the respective perspectives can then be used to identify the reifying tendencies in each of them. For Levine, this is how **reification may be “checked at the source”** and this is how a “critically reflexive moment might thus be rendered sustainable” (p. 103). It is in this sense that Levine's approach is not really post-foundational but, rather, an attempt to “balance foundationalisms against one another” (p. 14). There are strong parallels here with arguments advanced by assemblage thinking and complexity theory—links that could have been explored in more detail.

**Threats are real and the process of identifying them is valuable**

Jeffry A. **Frieden 5**, professor of government at Harvard University; and David A. Lake, professor of political science at the University of California, San Diego, July 2005, “International Relations as a Social Science: Rigor and Relevance,” Annals of the American Academy of Political and Social Science, No. 600, p. 136-156

International relationists have long been involved in foreign policy debates. The pages of thoughtful journals of opinion like Foreign Affairs or Foreign Policy are often filled by academics writing for broad audiences. And professors, of course, have frequently engaged in government service. There will always be a need for policy-relevant expertise. Through a lifetime of study, even the most theoretically inclined academics accumulate substantial country- or policy-specific knowledge that can supplement that possessed by those in government. Universities are repositories of country and policy experts “on call” to buttress hard-pressed policy makers confronted with crises in countries or over issues for which they l ack immediate knowledge. Yet in nearly all cases, the academics involved in past policy debates have acted as individuals, not as the embodiments of the accumulated knowledge of scholars of international relations. Rather, the real test of relevance is what the discipline provides in the way of approaches, theories, and analytical tools that can be marshaled to **explain why events happen** and **what can be done to alter the course of future events**. International Relations makes its biggest contribution—is most relevant—when it is **most “scientific**.” These contributions are of two types: well-developed, coherent, and convincing theoretical arguments for which there is **systematic empirical support**; and empirical evidence for a regularity, “law,” or other relationship that appears to hold beyond a reasonable doubt. **Only when International Relations brings science to the discussion** does it have **anything of enduring value to offer**, beyond well-informed opinion. And observers, policy makers, and journalists should be most influenced when scholars have something that is of enduring value to say. **Theory is not a substitute for policy expertise**, problem-solving abilities, or political experience; nor does it trump political struggle. Theories of international relations aim to capture general features of events and processes in ways that highlight their principal causes. Successful decision makers typically understand these forces intuitively, at least in the instances with which they are familiar. But explicit attention to scientific rigor can provide a degree of generality and clarity that might not be obvious even to experienced policy makers. Theory can also discipline the thinking of policy makers so that **sloppy** or **wishful thinking does not lead them astray**. The scientific study of international politics provides a logical and empirical check on attempts to draw inferences from the ideas or experiences of those who may be too close to the events they wish to influence. It would be a foolish policy maker who relied only on the scientific approach for guidance, but well-specified theories and tests of international relations are an essential part of a sound policy-making process.

**Security threats don’t cause militarization**

Vera **Zimmerman 15** has served as a research analyst at the Hudson Institute. Her scholarship has been published in World Politics Review. The author is currently a graduate student of political science in the Department of Public and International Affairs at George Mason University. Article Title “Book Review: The Theater of Operations: National Security Affect from the Cold War to the War on Terror. By Joseph Masco Durham and London: Duke University Press” – May 23rd, 2015 – Modified for potentially objectionable language - available at: https://verair.wordpress.com/2015/05/23/the-theater-of-operations-national-security-affect-from-the-cold-war-to-the-war-on-terror-by-joseph-masco-durham-and-london-duke-university-press-2014/

In chapter four, “Biosecurity Noir: WMDs in a World without Borders,” Masco singles out concrete evidence of the amplification of the invisible biothreat triggered by the receipt of a few anthrax letters in 2001 to support his argument about the made-up ambiguous link to WMD. The author highlights that by proliferating visions (depictions) of catastrophic danger, biosecurity created a militarized response **of global preemption** in the name of domestic defense. Masco argues his case well and sharply, providing compelling evidence, but his interpretations of evidence at times seems exaggerated and biased. Though Masco does not deny the existence of the real threats, his recognition of them is too brief, while consideration of an alternative view is rather weak. He acknowledges that terrorist violence is not fictitious but insists that for the most part the United States inflated threats and politically exploited potential danger to declare and maintain the state of emergency. Such focus on the amplification of threats seems to suggest that for the most part the threats are not that real. Masco suggests that the link between terrorists and WMD is mainly inflated. Yet there is a real global concern about 2,000 tons of highly-radioactive nuclear materials being stored in poorly secured civilian locations around the world. The book never mentions the threat of a dirty bomb, which today is viewed as a more likely occurrence than an atomic bomb explosion. The IAEA cites a hundred reported thefts of nuclear materials on average each year. There is a good chance terrorists can get their hands on enough nuclear materials to produce a dirty bomb. The United States meets these challenges with increased international cooperation. Masco’s main argument that “the US is no longer constrained by territorial limits” is **exaggerated**. The only two cases cited when the United States appeared unconstrained were the invasion of Iraq. Though the invasion of Iraq was opposed by some U.S. allies (France, Germany, and New Zealand), it was still a combined force coalition from the U.S., the UK, Australia, and Poland. The United States does not have an unrestricted reach as Masco wants to depict. It is constrained by sovereignty and territorial integrity of other stable states. The unstable nuclear regimes in **North Korea** and **Iran** present that existential nuclear threat to the U.S. described by Masco, but the United States is in **no rush to invade these countries**. According to the anticipatory and preemptive logic Masco prescribes to the United States, it could have already invaded those states to prevent the disaster. Another limitation of his argument is that he paints the nuclear and counterterror states as consistent through all the presidencies, thus, drawing all administrations under **a common denominator**. Under Obama, the counterterror state became a liberal democracy again. The ‘unrestrained’ theater of operation has shrunk **by ending the presence in Iraq and withdrawing from Afghanistan**, even though our presence there **could have been extended based on the preemption logic**. Obama recognized the faults of the Bush administration in acting unilaterally, scaled back stability operations, and emphasized sharing the costs and responsibilities of global leadership. The emergence of the real ISIL threat undermines the book’s core argument of threat amplification, the U.S. preemption logic of response, and unconstrained global reach. The U.S.-led global effort against ISIL amounts to more than 50 nations, which shows the unified nature of the fight. Masco asserts that U.S. superpower depends on the ability of the state to monopolize a discourse of danger, but he doesn’t discuss how the United States succeeded in doing that. Masco could have developed his argument by tracing how the United States was able to use its soft power to mobilize like-minded states to agree with U.S. hegemony on WOT. It will be interesting to trace the U.S. internalization of fear and terror. He could have examined how allies responded to U.S. domestic mobilization of its population and whether other states imitated U.S. emotional management projects to mobilize their own populations. This would boost his argument that the U.S. was able to project its power on the global scale. In Theatre of Operations, Masco makes a compelling argument about the creation of the unrestrained theater of operations via domestication of fear and terror carried over from the Cold War days. His anthropological study reveals the extent to which a democracy is willing to use fear to assure the core principle of the social contract, defined by Hobbes as the exchange of public obedience for collective security. A democracy that chooses to be preoccupied with security risks to forgo core democratic values resulting in the lack of transparency, restriction of free flow of information, and negligence of non-military threats—no less threatening than nuclear terrorism. Making criticism of U.S. actions the main focus of the book, however, Masco’s interpretations are not properly balanced and sometimes appear biased. Still, reading Masco’s insight of the purpose of U.S. actions in the post-9/11 context offers opportunities to think critically about the effects of 9/11 emotional reprogramming of society and state of emergencies in U.S. history.

\

**X-Risks OW**

**Extinction first**

**GPP 17** (Global Priorities Project, Future of Humanity Institute at the University of Oxford, Ministry for Foreign Affairs of Finland, “Existential Risk: Diplomacy and Governance,” Global Priorities Project, 2017, <https://www.fhi.ox.ac.uk/wp-content/uploads/Existential-Risks-2017-01-23.pdf>)

1.2. THE ETHICS OF EXISTENTIAL RISK In his book Reasons and Persons, Oxford philosopher Derek Parfit advanced an influential argument about the importance of avoiding extinction: I believe that if we destroy mankind, as we now can, this outcome will be much worse than most people think. Compare three outcomes: (1) Peace. (2) A nuclear war that kills 99% of the world’s existing population. (3) A nuclear war that kills 100%. (2) would be worse than (1), and (3) would be worse than (2). Which is the greater of these two differences? Most people believe that the greater difference is between (1) and (2). I believe that the difference between (2) and (3) is very much greater. ... The Earth will remain habitable for **at least another billion years**. Civilization began only a few thousand years ago. If we do not destroy mankind, these few thousand years may be only **a tiny fraction** of the whole of civilized human history. The difference between (2) and (3) may thus be the difference between this tiny fraction and all of the rest of this history. If we compare this possible history to a day, what has occurred so far is only a **fraction of a second**.65 In this argument, it seems that Parfit is assuming that the survivors of a nuclear war that kills 99% of the population would eventually be able to recover civilisation without long-term effect. As we have seen, this may not be a safe assumption – but for the purposes of this thought experiment, the point stands. What makes existential catastrophes especially bad is that they would “**destroy the future**,” as another Oxford philosopher, Nick Bostrom, puts it.66 This future could potentially be extremely long and full of flourishing, and would therefore have extremely large value. In standard risk analysis, when working out how to respond to risk, we work out the expected value of risk reduction, by weighing the probability that an action will prevent an adverse event against the severity of the event. Because the value of preventing existential catastrophe is so vast, even a tiny probability of prevention has huge expected value.67 Of course, there is persisting reasonable disagreement about ethics and there are a number of ways one might resist this conclusion.68 Therefore, it would be unjustified to be overconfident in Parfit and Bostrom’s argument. In some areas, government policy does give significant weight to future generations. For example, in assessing the risks of nuclear waste storage, governments have considered timeframes of thousands, hundreds of thousands, and even a million years.69 Justifications for this policy usually appeal to principles of intergenerational equity according to which future generations ought to get as much protection as current generations.70 Similarly, widely accepted norms of sustainable development require development that meets the needs of the current generation without compromising the ability of future generations to meet their own needs.71 However, when it comes to existential risk, it would seem that we fail to live up to principles of intergenerational equity. Existential catastrophe would not only give future generations less than the current generations; **it would give them nothing**. Indeed, reducing existential risk plausibly has a quite low cost for us in comparison with the huge expected value it has for future generations. In spite of this, relatively little is done to reduce existential risk. Unless we give up on norms of intergenerational equity, they give us a strong case for significantly increasing our efforts to reduce existential risks. 1.3. WHY EXISTENTIAL RISKS MAY BE SYSTEMATICALLY UNDERINVESTED IN, AND THE ROLE OF THE INTERNATIONAL COMMUNITY In spite of the importance of existential risk reduction, it probably receives less attention than is warranted. As a result, concerted international cooperation is required if we are to receive adequate protection from existential risks. 1.3.1. Why existential risks are likely to be underinvested in There are several reasons why existential risk reduction is likely to be underinvested in. Firstly, it is a global public good. Economic theory predicts that such goods tend to be underprovided. The benefits of existential risk reduction are widely and indivisibly dispersed around the globe from the countries responsible for taking action. Consequently, a country which reduces existential risk gains only a small portion of the benefits but bears the full brunt of the costs. Countries thus have strong incentives to free ride, receiving the benefits of risk reduction without contributing. As a result, too few do what is in the common interest. Secondly, as already suggested above, existential risk reduction is an intergenerational public good: most of the benefits are enjoyed by future generations who have no say in the political process. For these goods, the problem is temporal free riding: the current generation enjoys the benefits of inaction while future generations bear the costs. Thirdly, many existential risks, such as machine superintelligence, engineered pandemics, and solar geoengineering, pose an unprecedented and uncertain future threat. Consequently, it is hard to develop a satisfactory governance regime for them: there are few existing governance instruments which can be applied to these risks, and it is unclear what shape new instruments should take. In this way, our position with regard to these emerging risks is comparable to the one we faced when nuclear weapons first became available. Cognitive biases also lead people to underestimate existential risks. Since there have not been any catastrophes of this magnitude, these risks are not salient to politicians and the public.72 This is an example of the misapplication of the availability heuristic, a mental shortcut which assumes that something is important only if it can be readily recalled. Another cognitive bias affecting perceptions of existential risk is scope neglect. In a seminal 1992 study, three groups were asked how much they would be willing to pay to save 2,000, 20,000 or 200,000 birds from drowning in uncovered oil ponds. The groups answered $80, $78, and $88, respectively.73 In this case, the size of the benefits had little effect on the scale of the preferred response. People become numbed to the effect of saving lives when the numbers get too large. 74 Scope neglect is a particularly acute problem for existential risk because the numbers at stake are so large. Due to scope neglect, decision-makers are prone to treat existential risks in a similar way to problems which are less severe by many orders of magnitude. A wide range of other cognitive biases are likely to affect the evaluation of existential risks.75

**Predictions of high-magnitude impacts are accurate and valuable.**

**Gleditsch 12** (Kristian S. Gleditsch, Department of Government, University of Essex & Peace Research Institute Oslo, and Michael D. Ward, Department of Political Science, Duke, 2012, “Forecasting is difficult, especially about the future: Using contentious issues to forecast interstate disputes,” Journal of Peace Research, 50(1) 17–31, Sage Journals)  
There have been **remarkably few efforts** to generate global forecasts or risk profiles for interstate conflict. Moreover, the most prominent efforts to consider the predictive ability of models of interstate conflicts have based their research on models that were not actually proposed with forecasting in mind. A notable example here is Beck, King & Zeng (2000), who essentially adopt the so-called liberal peace model of Russett & Oneal (2001). Certainly, nothing akin to the Political Instability Task Force’s now annual projections (beginning with Gurr, Marshall & Khosla, 2000) exists for international conflicts. Perhaps not surprisingly, many observers are **very skeptical** of the ability of academic researchers to anticipate conflict between states, at least beyond very short time horizons. Research in recent decades has seen a large number of hypotheses generated to explain under what conditions militarized interstate conflict is more or less likely. This avenue of research has been primarily inspired by research on the so-called democratic peace, or the absence of conflict between democracies. Indeed, there are thousands of scholarly works mentioning the term militarized interstate dispute (MID), most of which use these data for some kind of empirical examination of a proposition about disputes. Yet, the evidence suggests that the ability of this body of work to forecast conflict out-of-sample is **decidedly disappointing**. Ward, Siverson & Cao (2007) found that most of the recent statistical studies of militarized interstate disputes in prominent political science and international relations journals were unable to predict the outbreak of a single dispute out-of-sample (see also Beck, King & Zeng, 2000). Many researchers have sought to improve on the ability to forecast militarized interstate conflict by turning to alternative statistical methods. Beck, King & Zeng (2000), for example, find that neural networks perform marginally better than generalized linear regression models in forecasting conflict from the same input factors.1 Changes in estimation methods or statistical techniques per se, however, have at best led only to limited improvements in out-of-sample predictive ability. Our argument is that **simply identifying inappropriate methods** as the key source of the problem in forecasting conflict may give us the **wrong diagnosis** and lead us down less productive avenues. A more fundamental problem is models that provide a poor basis for forecasting by disregarding the motives for conflict to arise, or by only considering motives in a relatively superficial manner. Models that have been proposed for research on the democratic peace, notably the work of Russett & Oneal (2001), are primarily intended to examine whether certain characteristics of liberal institutions, such as democracy and trade, make conflict on average less likely relative to baseline risks of conflict. Although these approaches may be appropriate for testing the original propositions of interest, they essentially ignore the **contentious issues** that might cause states to resort to violence and instead treat these contentious issues as exogenous features, typically hidden inside a so-called ‘black box’ of the baseline risk of conflict. Our own initial foray into out-of-sample prediction for a state-level model indicates that spatial information about other conflict events can help to improve forecasts (see Ward & Gleditsch, 2002). Although this allows predictions of conflict to be conditional on other observed events rather than treating each conflict as an independent observation, the approach still ignores the issues over which such conflicts may have arisen initially. We believe that greater attention to the **specific reasons for the occurrence of conflicts** and the incompatibilities that may generate the use of violence can help **improve our ability to forecast conflict.** Although we recognize that different models may be appropriate to evaluate particular propositions and to forecast events, in our view **the enterprise of prediction has great potential** for winnowing bad ideas out of theories on the causes of conflict and avoiding the problem of retrospective biases in conventional hypothesis testing on the data used to develop the hypotheses in the first place (see Ward, Greenhill & Bakke, 2010). In fairness, much of the existing work on the statistical modeling of conflict has bypassed motivation since it is genuinely difficult to establish what states fight over and what their possible motivation for fighting might be. Nevertheless, **the fact that something is difficult to evaluate does not mean that simply ignoring it is the best course of action**. Another tradition in research on conflict has sought to identify incompatibilities in terms of contentious issues, such as territorial or maritime claims (Diehl, 1992; Mansbach & Vasquez, 1981). Recent efforts to examine these propositions empirically have found considerable evidence that cases where such claims exist are more likely to see militarized activities (Hensel, 2001; Hensel et al., 2008; Hensel & Mitchell, 2010). Even so, at present this line of research has primarily engaged in testing hypotheses about whether coefficient estimates are significantly different from 0 or the in-sample post-diction of conflicts, and has not yet examined if information on contentious issues may be helpful for forecasting dyadic conflict out-of-sample. Here we explicitly consider whether taking into account information on contentious issues and conflict management can help improve on forecasting interstate conflict and our understanding of conflict dynamics. Although we focus on statistical approaches to interstate conflict in this article, many of our arguments also apply to problems in traditional theories of conflict and qualitative approaches to prediction or anticipating political events (see Tetlock, 2005). Traditional theories of interstate conflict tend to focus on **structural features** presumed to influence the opportunities for conflict such as the distribution of power in the international system or relative balance of power (see e.g. Waltz, 1979). These theories display little interest in the specific incompatibilities that may motivate the use of violence. However, structural factors rarely change rapidly, but violent conflict tends to be episodic, and hence cannot be adequately explained merely by reference to permissive conditions (see Fearon, 1995). Likewise, our core argument applies to studies of civil war, which tend to emphasize opportunities for conflict rather than motivations for conflict (see Cederman, Weidmann & Gleditsch, 2011), and where evidence for the predictive ability of existing statistical efforts seems similarly disappointing (see Ward, Greenhill & Bakke, 2010). Many political and area study experts, typically using informal methods for deriving predictions, often have strong confidence in their ability to forecast events. However, the comprehensive series of studies by Tetlock (2005), who asked experts to rate a series of outcomes which could then be compared against the historical record, provide little support for the forecasting ability of political experts.

**Prioritize existential risk prevention---it encompasses AND outweighs other threats.**

Dennis **Pamlin &** Stuart **Armstrong 15**, Dennis Pamlin, Executive Project Manager Global Risks, Global Challenges Foundation, and Stuart Armstrong, James Martin Research Fellow, Future of Humanity Institute, Oxford Martin School, University of Oxford, February 2015, “Global Challenges: 12 Risks that threaten human civilization: The case for a new risk category,” Global Challenges Foundation, p.30-93, https://api.globalchallenges.org/static/wp-content/uploads/12-Risks-with-infinite-impact.pdf

2. Risks with infinite impact: A new category of risks “Most risk management is really just advanced contingency planning and disciplining yourself to realise that, given enough time, very **low probability events not only can happen, but they absolutely will happen**.” Lloyd Blankfein, Goldman Sachs CEO, July 2013 1 **Risk = Probability × Impact** Impacts where civilisation collapses to a state of **great suffering** and **do not recover**, or a situation where all human life end, are defined as **infinite** as the result is **irreversible** and **lasts forever**. A new group of global risks This is a report about a limited number of global risks – that can be identified through a scientific and transparent process – with impacts of a magnitude that pose a threat to human civilisation, or even possibly to all human life. With such a focus it may surprise some readers to find that the report’s essential aim is to inspire action and dialogue as well as an increased use of the methodologies used for risk assessment. The real focus is not on the almost unimaginable impacts of the risks the report outlines. Its fundamental purpose is to encourage global collaboration and to use this new category of risk as a driver for innovation. The idea that we face a number of global challenges threatening the very basis of our civilisation at the beginning of the 21st century is well accepted in the scientific community, and is studied at a number of leading universities.2 But there is still no coordinated approach to address this group of challenges and turn them into opportunities for a new generation of global cooperation and the creation of a global governance system capable of addressing the greatest challenges of our time. This report has, to the best of our knowledge, created the first science-based list of global risks with a potentially infinite impact and has made the first attempt to provide an initial overview of the uncertainties related to these risks as well as rough quantifications for the probabilities of these impacts. What is risk? Risk is the potential of losing something of value, weighed against the potential to gain something of value. Every day we make different kinds of risk assessments, in more or less rational ways, when we weigh different options against each other. The basic idea of risk is that an uncertainty exists regarding the outcome and that we must find a way to take the best possible decision based on our understanding of this uncertainty.3 To calculate risk the probability of an outcome is often multiplied by the impact. The impact is in most cases measured in economic terms, but it can also be measured in anything we want to avoid, such as suffering. At the heart of a risk assessment is a probability distribution, often described by a probability density function4; see figure X for a graphic illustration. The slightly tilted bell curve is a common probability distribution, but the shape differs and in reality is seldom as smooth as the example. The total area under the curve always represents 100 percent, i.e. all the possible outcomes fit under the curve. In this case (A) represents the most probable impact. With a much lower probability it will be a close to zero impact, illustrated by (B). In the same way as in case B there is also a low probability that the situation will be very significant, illustrated by (C). Figure 1: Probability density function [FIGURE 1 OMITTED] The impacts (A), (B) and (C) all belong to the same category, ~~normal~~ [common] impacts: the impacts may be more or less serious, but they can be dealt with within the current system. The impacts in this report are however of a special kind. These are impacts where everything will be lost and the situation will not be reversible, i.e challenges with potentially infinite impact. In insurance and finance this kind of risk is called “risk of ruin”, an impact where all capital is lost.5 This impact is however only infinite for the company that is losing the money. From society’s perspective, that is not a special category of risk. In this report the focus is on the “risk of ruin” on a global scale and on a human level, in the worst case this is when we risk the extinction of our own species. On a probability curve the impacts in this report are usually at the very far right with a relatively low probability compared with other impacts, illustrated by (D) in Figure 2. Often they are so far out on the tail of the curve that they are not even included in studies. For each risk in this report the probability of an infinite impact is very low compared to the most likely outcome. Some studies even indicate that not all risks in this report can result in an infinite impact. But a significant number of **peer-reviewed** reports indicate that those impacts **not only can happen**, but that their **probability is increasing** due to **unsustainable trends**. The assumption for this report is that by creating a better understanding of our scientific knowledge regarding risks with a potentially infinite impact, we can inspire initiatives that can turn these risks into drivers for innovation. Not only could a better understanding of the unique magnitude of these risks help address the risks we face, it could also help to create a path towards more sustainable development. The group of global risks discussed in this report are so different from most of the challenges we face that they are hard to comprehend. But that is also why they can help us to build the collaboration we need and drive the development of further solutions that benefit both people and the planet. As noted above, none of the risks in this report is likely to result directly in an infinite impact, and some are probably even physically incapable of doing so. But all are so significant that they could reach a threshold impact able to create social and ecological instability that could trigger a process which could lead to an infinite impact. For several reasons the potentially infinite impacts of the risks in this report are not as well known as they should be. One reason is the way that extreme impacts are often masked by most of the theories and models used by governments and business today. For example, the probability of extreme impacts is often below what is included in studies and strategies. The tendency to exclude impacts below a probability of five percent is one reason for the **relative “invisibility”** of **infinite impacts**. The almost standard use of a 95% confidence interval is one reason why low-probability high-impact events are often ignored.6 Figure 2: Probability density function with tail highlighted [FIGURE 2 OMITTED] Climate change is a good example, where almost all of the focus is on the most likely scenarios and there are few studies that include the low-probability high-impact scenarios. In most reports about climate impacts, the impacts caused by warming beyond five or six degrees Celsius are even omitted from tables and graphs even though the IPCC’s own research indicates that the probability of these impacts are often between one and five percent, and sometimes even higher.7 Other aspects that contribute to this relative invisibility include the fact that extreme impacts are difficult to translate into monetary terms, they have a global scope, and they often require a time-horizon of a century or more. They cannot be understood simply by linear extrapolation of current trends, and they lack historical precedents. There is also the fact that the measures required to significantly reduce the probability of infinite impacts will be radical compared to a business-as-usual scenario with a focus on incremental changes. The exact probability of a specific impact is difficult or impossible to estimate.8 However, the important thing is to establish the current magnitude of the probabilities and compare them with the probabilities for such impacts we cannot accept. A failure to provide any estimate for these risks often results in strategies and priorities defined as though the probability of a totally unacceptable outcome is zero. An approximate number for a best estimate also makes it easier to understand that a great uncertainty means the actual probability can be both much higher and much lower than the best estimate. It should also be stressed that uncertainty is not a weakness in science; it always exists in scientific work. It is a systematic way of understanding the limitations of the methodology, data, etc.9 Uncertainty is not a reason to wait to take action if the impacts are serious. Increased uncertainty is something that risk experts, e.g. insurance experts and security policy experts, interpret as a signal for action. A contrasting challenge is that our cultural references to the threat of infinite impacts have been dominated throughout history by religious groups seeking to scare society without any scientific backing, often as a way to **discipline people** and implement unpopular measures. It should not have to be said, but this report is obviously **fundamentally different** as it focuses on **scientific evidence** from **peer-reviewed sources**. Infinite impact The concept infinite impact refers to two aspects in particular; the terminology is not meant to imply a literally infinite impact (with all the mathematical subtleties that would imply) but to serve as a reminder that these risks are of a different nature. Ethical These are impacts that threaten the very **survival of humanity and life on Earth** – and therefore can be seen as being infinitely negative from an **ethical perspective**. **No positive gain can outweigh even a small probability** for an infinite negative impact. Such risks require society to ensure that we eliminate these risks by reducing the impact below an infinite impact as a **top priority**, or at least do everything we can to reduce the probability of these risks. As some of these risks are impossible to eliminate today it is also important to discuss what probability can right now be accepted for risks with a possible infinite impact. Economic Infinite impacts are beyond what most traditional economic models today are able to cope with. The impacts are irreversible in the most fundamental way, so tools like cost-benefit assessment seldom make sense. To use discounting that makes infinite impacts (which could take place 100 years or more from now and affect all future generations) close to invisible in economic assessments, is another example of a challenge with current tools. So while tools like cost-benefit models and discounting can help us in some areas, they are seldom applicable in the context of infinite impacts. New tools are needed to guide the global economy in an age of potential infinite impacts. See chapter 2.2.2 for a more detailed iscussion. Roulette and Russian roulette When probability and normal risks are discussed the example of a casino and roulette is often used. You bet something, then spin the wheel and with a certain probability you win or lose. You can use different odds to discuss different kinds of risk taking. These kinds of thought experiment can be very useful, but when it comes to infinite risks these gaming analogies become problematic. For infinite impact a more appropriate analogy is probably Russian roulette. But instead of “normal” Russian roulette where you only bet your own life you are now also betting everyone you know and everyone you don’t know. Everyone alive will die if you lose. There will be no second chance for anyone as there will be no future generations; humanity will end with your loss. What probability would you accept for different sums of money if you played this version of Russian roulette? Most people would say that it is stupid and – no matter how low the probability is and no matter how big the potential win is – this kind of game should not be played, as it is unethical. Many would also say that no person should be allowed to make such a judgment, as those who are affected do not have a say. You could add that most of those who will lose from it cannot say anything as they are not born and will never exist if you lose. The difference between ordinary roulette and “allhumanity Russian roulette” is one way of illustrating the difference in nature between a “normal” risk that is reversible, and a risk with an infinite impact. An additional challenge in acknowledging the risks outlined in this report is that many of the traditional risks including wars and violence have decreased, even though it might not always looks that way in media.10 So a significant number of experts today spend a substantial amount of time trying to explain that much of what is discussed as dangerous trends might not be as dangerous as we think. For policy makers listening only to experts in traditional risk areas it is therefore easy to get the impression that global risks are becoming less of a problem. The **chain of events** that could result in **infinite impacts** in this report also **differ** from most of the traditional risks, as most of them are **not** triggered by **wilful acts**, but **accidents**/mistakes. Even the probabilities related to nuclear war in this report are to a large degree related to inadvertent escalation. As many of the tools to analyse and address risks have been developed to protect nations and states from attacks, risks involving accidents tend to get less attention. This report emphasises the need for an open and democratic process in addressing global challenges with potentially infinite impact. Hence, this is a scientifically based invitation to discuss how we as a global community can address what could be considered the greatest challenges of our time. The difficulty for individual scientists to communicate a scientific risk approach should however not be underestimated. Scientists who today talk about low-probability impacts, that are serious but still far from infinite, are often accused of pessimism and scaremongering, even if they do nothing but highlight scientific findings.11 To highlight infinite impacts with even lower probability can therefore be something that a scientist who cares about his/her reputation would want to avoid. In the media it is still common to contrast the most probable climate impact with the probability that nothing, or almost nothing, will happen. The fact that almost nothing could happen is not wrong in most cases, but it is unscientific and dangerous if different levels of probability are presented as equal. The tendency to compare the most probable climate impact with the possibility of a low or no impact also results in a situation where low-probability high-impact outcomes are often totally ignored. An honest and scientific approach is to, whenever possible, present the whole probability distribution and pay special attention to unacceptable outcomes. The fact that we have challenges that with some probability might be infinite and therefore fundamentally irreversible is difficult to comprehend, and **physiologically** they are something our brains are poorly equipped to respond to, according to evolutionary psychologists.12 It is hard for us as individuals to grasp that humanity for the first time in its history now has the capacity to create such catastrophic outcomes. Professor Marianne Frankenhaeuser, former head of the psychology division, Karolinska Institute, Stockholm, put it this way: “Part of the answer is to be found in psychological defence mechanisms. The nuclear threat is collectively denied, because to face it would force us to face some aspects of the world’s situation which we do not want to recognise.” 13 This **psychological** denial may be one reason why there is a tendency among some stakeholders to confuse “being optimistic” with denying what science is telling us, and ignoring parts of the probability curve.14 Ignoring the fact that there is strong scientific evidence for serious impacts in different areas, and focusing only on selected sources which suggest that the problem may not be so serious, is not optimistic. It is both unscientific and dangerous.15 A scientific approach requires us to base our decisions on the whole probability distribution. Whether it is possible to address the challenge or not is the area where optimism and pessimism can make people look at the same set of data and come to different conclusions. Two things are important to keep in mind: first, that there is always a probability distribution when it comes to risk; second, that there are two different kinds of impacts that are of interest for this report. The probability distribution can have different shapes but in simplified cases the shape tends to look like a slightly modified clock (remember figure 1). In the media it can sound as though experts argue whether an impact, for example a climate impact or a pandemic, will be dangerous or not. But what serious experts discuss is the probability of different oucomes. They can disagree on the shape of the curve or what curves should be studied, but not that a probability curve exists. With climate change this includes discussions about how sensitive the climate is, how much greenhouse gas will be emitted, and what impacts that different warmings will result in. Just as it is important not to ignore challenges with potentially infinite impacts, it is also important not to use them to scare people. Dramatic images and strong language are best avoided whenever possible, as this group of risks require sophisticated strategies that benefit from rational arguments. Throughout history we have seen too many examples when threats of danger have been damagingly used to undermine important values. The history of infinite impacts: The LA-602 document The understanding of infinite impacts is very recent compared with most of our institutions and laws. It is only 70 years ago that Edward Teller, one of the greatest physicists of his time, with his back-of-the-envelope calculations, produced results that differed drastically from all that had gone before. His calculations indicated that the explosion of a nuclear bomb – a creation of some of the brightest minds on the planet, including Teller himself – could result in a chain reaction so powerful that it would ignite the world’s atmosphere, thereby ending human life on Earth.16 Robert Oppenheimer, who led the Manhattan Project to develop the nuclear bomb, halted the project to see whether Teller’s calculations were correct.17 The resulting document, LA- 602: Ignition of the Atmosphere with Nuclear Bombs, concluded that Teller was wrong, But the sheer complexity drove them to end their assessment by writing that “further work on the subject [is] highly desirable”.18 The LA-602 document can be seen as the first scientific global risk report addressing a category of risks where the worst possible impact in all practical senses is infinite.19 Since the atomic bomb more challenges have emerged with potentially infinite impact. Allmost all of these new challenges are linked to the increased knowledge, economic and technical development that has brought so many benefits. For example, climate change is the result of the industrial revolution and development that was, and still is, based heavily on fossil fuel. The increased potential for global pandemics is the result of an integrated global economy where goods and services move quickly around the world, combined with rapid urbanisation and high population density. In parallel with the increased number of risks with possible infinite impact, our capacity to analyse and solve them has greatly increased too. Science and technology today provides us with knowledge and tools that can radically reduce the risks that historically have been behind major extinctions, such as pandemics and asteroids. Recent challenges like climate change, and emerging challenges like synthetic biology and nanotechnology, can to a large degree be addressed by smart use of new technologies, new lifestyles and institutional structures. It will be hard as it will require collaboration of a kind that we have not seen before. It will also require us to create systems that can deal with the problems before they occur. The fact that the same knowledge and tools can be both a problem and a solution is important to understand in order to avoid polarisation. Within a few decades, or even sooner, many of the tools that can help us solve the global challenges of today will come from fields likely to provide us with the most powerful instruments we have ever had – resulting in their own sets of challenges. Synthetic biology, nanotechnology and artificial intelligence (AI) are all rapidly evolving fields with great potential. They may help solve many of today’s main challenges or, if not guided in a benign direction, may result in catastrophic outcomes. The point of departure of this report is the fact that we now have the knowledge, economic resources and technological ability to reduce most of the greatest risks of our time. Conversely, the infinite impacts we face are almost all unintended results of human ingenuity. The reason we are in this situation is that we have made progress in many areas without addressing unintended low-probability high-impact consequences. Creating innovative and resilient systems rather than simply managing risk would let us focus more on opportunities. But the resilience needed require moving away from legacy systems is likely to be disruptive, so an open and transparent discussion is needed regarding the transformative solutions required. Figure 3: Probability density function with tail and threshold highlighted [FIGURE 3 OMITTED] 2.1 Report structure The first part of the report is an introduction where the global risks with potential infinite impact are introduced and defined. This part also includes the methodology for selecting these risks, and presents the twelve risks that meet this definition. Four goals of the report are also presented, under the headings “acknowledge”, “inspire”, “connect” and “deliver”. The second part is an overview of the twelve global risks and key events that illustrate some of the work around the world to address them. For each challenge five important factors that influence the probability or impact are also listed. The risks are divided into four different categories depending on their characteristics. “Current challenges” is the first category and includes the risks that currently threaten humanity due to our economic and technological development - extreme climate change, for example, which depends on how much greenhouse gas we emit. “Exogenic challenges” includes risks where the basic probability of an event is beyond human control, but where the probability and magnitude of the impact can be influenced - asteroid impacts, for example, where the asteroids’ paths are beyond human control but an impact can be moderated by either changing the direction of the asteroid or preparing for an impact. “Emerging challenges” includes areas where technological development and scientific assessment indicate that they could both be a very important contribution to human welfare and help reduce the risks associated with current challenges, but could also result in new infinite impacts.20 AI, nanotechnology and synthetic biology are examples. “Global policy challenge” is a different kind of risk. It is a probable threat arising from future global governance as it resorts to destructive policies, possibly in response to the other challenges listed above. The third part of the report discusses the relationship between the different risks. Action to reduce one risk can increase another, unless their possible links are understood. Many solutions are also able to address multiple risks, so there are significant benefits from understanding how one relates to others. Investigating these correlations could be a start, but correlation is a linear measure and non-linear techniques may be more helpful for assessing the aggregate risk. The fourth part is an overview, the first ever to our knowledge, of the uncertainties and probabilities of global risks with potentially infinite impacts. The numbers are only rough estimates and are meant to be a first step in a dialogue where methodologies are developed and estimates refined. The fifth part presents some of the most important underlying trends that influence the global challenges, which often build up slowly until they reach a threshold and very rapid changes ensue. The sixth and final part presents an overview of possible ways forward. 2.2 Goals Goal 1: Acknowledge That key stakeholders, influencing global challenges, acknowledge the existence of the category of risks that could result in infinite impact. They should also recognice that the list of risks that belong to this category should be revised as new technologies are developed and our knowledge increases. Regardless of the risks included, the category should be given special attention in all processes and decisions of relevance. The report also seeks to demonstrate to all key stakeholders that we have the capacity to reduce, or even eliminate, most of the risks in this category. Establish a category of risks with potentially infinite impact. Before anything significant can happen regarding global risks with potentially infinite impacts, their existence must be acknowledged. Rapid technological development and economic growth have delivered unprecedented material welfare to billions of people in a veritable tide of utopias.21 But we now face the possibility that even tools created with the best of intentions can have a darker side too, a side that may threaten human civilisation, and conceivably the continuation of human life. This is what all decision-makers need to recognise. Rather than succumbing to terror, we need to acknowledge that we can let the prospect inspire and drive us forward. Goal 2: Inspire That policy makers inspire action by explaining how the probabilities and impacts can be reduced and turned into opportunities. Concrete examples of initiatives should be communicated in different networks in order to create ripple effects, with the long-term goal that all key stakeholders should be inspired to turn these risks into opportunities for positive action. Show concrete action that is taking place today. This report seeks to show that it is not only possible to contribute to reducing these risks, but that it is perhaps the most important thing anyone can spend their time on. It does so by combining information about the risks with information about individuals and groups who has made a significant contribution by turning challenges into opportunities. By highlighting concrete examples the report hopes to inspire a new generation of leaders. Goal 3: Connect That leaders in different sectors connect with each other to encourage collaboration. A specific focus on financial and security policy where significant risks combine to demand action beyond the incremental is required. Support new meetings between interested stakeholders. The nature of these risks spans countries and continents; they require action by governments and politicians, but also by companies, academics, NGOs, and many other groups. The magnitude of the possible impacts requires not only leaders to act but above all new models for global cooperation and decision-making to ensure delivery. The need for political leadership is therefore crucial. Even with those risks where many groups are involved, such as climate change and pandemics, very few today address the possibility of infinite impact aspects. Even fewer groups address the links between the different risks. There is also a need to connect different levels of work, so that local, regional, national and international efforts can support each other when it comes to risks with potentially infinite impacts. Goal 4: Deliver That concrete strategies are developed that allow key stakeholders to identify, quantify and address global challenges as well as gather support for concrete steps towards a wellfunctioning global governance system. This would include tools and initiatives that can help identify, quantify and reduce risks with potentially infinite impacts. Identify and implement strategies and initiatives. Reports can acknowledge, inspire and connect, but only people can deliver actual results. The main focus of the report is to show that actual initiatives need to be taken that deliver actual results. Only when the probability of an infinite impact becomes acceptably low, very close to zero, and/or when the maximum impact is significantly reduced, should we talk about real progress. In order to deliver results it is important to remember that global governance to tackle these risks is the way we organise society in order to address our greatest challenges. It is not a question of establishing a “world government”, it is about the way we organise ourselves on all levels, from the local to the global. The report is a first step and should be seen as an invitation to all responsible parties that can affect the probability and impact of risks with potentially infinite impacts. But its success will ultimately be measured only on how it contributes to concrete results. 2.3 Global challenges and infinite impact This chapter first introduces the concept of infinite impact. It then describes the methodology used to identify challenges with an infinite impact. It then presents risks with potentially infinite impact that the methodology results in. 2.3.1 Definition of infinite impact The specific criterion for including a risk in this report is that well-sourced science shows the challenge can have the following consequences: 22 1. Infinite impact: When civilisation collapses to a state of great suffering and does not recover, or a situation where all human life ends. The existence of such threats is well attested by science.23 2. Infinite impact threshold – an impact that can trigger a chain of events that could result first in a civilisation collapse, and then later result in an infinite impact. Such thresholds are especially important to recognise in a complex and interconnected society where resilience is decreasing.24 A collapse of civilisation is defined as a drastic decrease in human population size and political/economic/social complexity, globally for an extended time.25 The above definition means the list of challenges is not static. When new challenges emerge, or current ones fade away, the list will change. An additional criterion for including risks in this report is “human influence”. Only risks where humans can influence either the probability, the impact, or both, are included. For most risks both impact and probability can be affected, for example with nuclear war, where the number/size of weapons influences the impact and tensions between countries affects the probability. Other risks, such as a supervolcano, are included as it is possible to affect the impact through various mitigation methods, even if we currently cannot affect the probability. Risks that are susceptible to human influence are indirectly linked, because efforts to address one of them may increase or decrease the likelihood of another. 2.3.2 Why use “infinite impact” as a concept? The concept of infinity was chosen as it reflects many of the challenges, especially in economic theory, to addressing these risks as well as the need to question much of our current way of thinking. The concept of a category of risks based on their extreme impact is meant to provide a tool to distinguish one particular kind of risk from others. The benefit of this new concept should be assessed based on two things. First, does the category exist, and second, is the concept helpful in addressing these risks? The report has found ample evidence that there are risks with an impact that can end human civilisation and even all human life. The report further concludes that a new category of risk is not only meaningful but also timely. We live in a society where global risks with potentially infinite impacts increase in both number and probability according to multiple studies. Looking ahead, many emerging technologies which will certainly provide beneficial results, might also result in an increased probability of infinite impacts.26 Over the last few years a greater understanding of low probability or unknown probability events has helped more people to understand the importance of looking beyond the most probable scenarios. Concepts like “black swans” and “perfect storms” are now part of mainstream policy and business language.27 Greater understanding of the technology and science of complex systems has also resulted in a new understanding of potentially disruptive events. Humans now have such an impact on the planet that the term “the anthropocene” is being used, even by mainstream media like The Economist.28 The term was introduced in the 90s by the Nobel Prize winner Paul Crutzen to describe how humans are now the dominant force changing the Earth’s ecosystems.29 The idea to establish a well defined category of risks that focus on risks with a potentially infinite impact that can be used as a practical tool by policy makers is partly inspired by Nick Bostrom’s philosophical work and his introduction of a risk taxonomy that includes an academic category called “existential risks”.30 Introducing a category with risks that have a potentially infinite impact is not meant to be a mathematical definition; infinity is a thorny mathematical concept and nothing in reality can be infinite.31 It is meant to illustrate a singularity, when humanity is threatened, when many of the tools used to approach most challenges today become problematic, meaningless, or even counterproductive. The concept of an infinite impact highlights a unique situation where humanity itself is threatened and the very idea of value and price collapses from a human perspective, as the price of the last humans also can be seen to be infinite. This is not to say that those traditional tools cannot still be useful, but with infinite impacts we need to add an additional set of analytical tools. Life Value The following estimates have been applied to the value of life in the US. The estimates are either for one year of additional life or for the statistical value of a single life. – $50,000 per year of quality life (international standard most private and government-run health insurance plans worldwide use to determine whether to cover a new medical procedure) – $129,000 per year of quality life (based on analysis of kidney dialysis procedures by Stefanos Zenios and colleagues at Stanford Graduate School of Business) – $7.4 million (Environmental Protection Agency) – $7.9 million (Food and Drug Administration) – $6 million (Transportation Department) – $28 million (Richard Posner based on the willingness to pay for avoiding a plane crash) Source: Wikipedia: Value of life http://en.wikipedia.org/wiki/Value\_of\_life US EPA: Frequently Asked Questions on Mortality Risk Valuation http://yosemite.epa.gov/EE%5Cepa%5Ceed.nsf/webpages/MortalityRiskValuation.html Posner, Richard A. Catastrophe: risk and response. Oxford University Press, 2004 Some of the risks, including nuclear war, climate change and pandemics, are often included in current risk overviews, but in many cases their possible infinite impacts are excluded. The impacts which are included are in most cases still very serious, but only the more probable parts of the probability distributions are included, and the last part of the long tail – where the infinite impact is found – is excluded.32 Most risk reports do not differentiate between challenges with a limited impact and those with a potential for infinite impact. This is dangerous, as it can mean resources are spent in ways that increase the probability of an infinite impact. Ethical aspects of infinite impact The basic ethical aspect of infinite impact is this: a very small group alive today can take decisions that will fundamentally affect all future generations. “All future generations” is not a concept that is often discussed, and for good reason. All through human history we have had no tools with a measurable global impact for more than a few generations. Only in the last few decades has our potential impact reached a level where all future generations can be affected, for the simple reason that we now have the technological capacity to end human civilisation. If we count human history from the time when we began to practice settled agriculture, that gives us about 12,000 years.33 If we make a moderate assumption that humanity will live for at least 50 million more years34 our 12,000-year history so far represents 1/4200, or 0.024%, of our potential history. So our generation has the option of risking everything and annulling 99.976% of our potential history. Comparing 0.024% with the days of a person living to 100 years from the day of conception, this would equal less than nine days and is the first stage of human embryogenesis, the germinal stage.35 Two additional arguments to treat potentially infinite impacts as a separate category are: 36 1. **An approach to infinite impacts cannot be one of trial-and-error**, because there is no opportunity to learn from errors. The reactive approach – see what happens, limit damage, and learn from experience – is unworkable. Instead society must be proactive. This requires foresight to foresee new types of threat and willingness to take decisive preventative action and to bear the costs (**moral** and economic) of such actions. 2. We cannot necessarily rely on the institutions, morality, social attitudes or national security policies that developed from our experience of other sorts of risk. Infinite impacts are in a different category. Institutions and individuals may find it hard to take these risks seriously simply because they lie outside our experience. Our collective fear-response will probably be ill-calibrated to the magnitude of threat. Economic aspects of infinite impact and discounting In today’s society a monetary value is sometimes ascribed to human life. Some experts use this method to estimate risk by assigning a monetary value to human extinction.37 We have to remember that the monetary values placed on a human life in most cases are not meant to suggest that we have actually assigned a specific value to a life. Assigning a value to a human life is a tool used in a society with a limited supply of resources or infrastructure (ambulances, perhaps) or skills. In such a society it is impossible to save every life, so some trade-off must be made.38 The US Environmental Protection Agency explains its use like this: “The EPA does not place a dollar value on individual lives. Rather, when conducting a benefit-cost analysis of new environmental policies, the Agency uses estimates of how much people are willing to pay for small reductions in their risks of dying from adverse health conditions that may be caused by environmental pollution.” 39 The fact that monetary values for human lives can help to define priorities when it comes to smaller risks does not mean that they are suitable for quite different uses. Applying a monetary value to the whole human race makes little sense to most people, and from an economic perspective it makes no sense. Money helps us to prioritise, but with no humans there would be no economy and no need for priorities. Ignoring, or discounting, future generations is actually the only way to avoid astronomical numbers for impacts that may seriously affect every generation to come. In Catastrophe: Risk and Response, Richard Posner provides a cost estimate, based on the assumption that a human life is worth $50,000, resulting in a $300 tn cost for the whole of humanity, assuming a population of six billion. He then doubles the population number to include the value of all future generations, ending up with $600 tn, while acknowledging that “without discounting, the present value of the benefits of risk-avoidance measures would often approach infinity for the type of catastrophic risk with which this book is concerned.” 40 Discounting for risks that include the possibility of an infinite impact differs from risk discounting for less serious impacts. For example the Stern Review41 prompted a discussion between its chief author, Nicholas Stern, and William Nordhaus,42 each of whom argued for different discount levels using different arguments. But neither discussed a possible infinite climate impact. An overview of the discussion by David Evans of Oxford Brookes University highlighted some of the differing assumptions.43 Two things make infinite impacts special from a discounting perspective. First, there is no way that future generations can compensate for the impact, as they will not exist. Second, the impact is something that is beyond an individual preference, as society will no longer exist. Discounting is undertaken to allocate resources in the most productive way. In cases that do not include infinite impacts, discounting “reflects the fact that there are many high-yield investments that would improve the quality of life for future generations. The discount rate should be set so that our investable funds are devoted to the most productive uses.” 44 When there is a potentially infinite impact, the focus is no longer on what investments have the best rate of return, it is about avoiding the ultimate end. While many economists shy away from infinite impacts, those exploring the potentially extreme impacts of global challenges often assume infinite numbers to make their point. Nordhaus for example writes that “the sum of undiscounted anxieties would be infinite (i.e. equal to 1 + 1 +1 + … = ∞). In this situation, most of us would dissolve in a sea of anxiety about all the things that could go wrong for distant generations from asteroids, wars, out-of-control robots, fat tails, smart dust and other disasters.” 45 It is interesting that Nordhaus himself provides very good graphs that show why the most important factor when determining actions is a possible threshold (see below Figure 4 and 5). Nordhaus was discussing climate change, but the role of thresholds is similar for most infinite impacts. The first figure is based on traditional economic approaches which assume that Nature has no thresholds; the second graph illustrates what happens with the curve when a threshold exists. As Nordhaus also notes, it is hard to establish thresholds, but if they are significant all other assumptions become secondary. The challenge that Nordhaus does not address, and which is important especially with climate change, is that thresholds become invisible in economic calculations if they occur far into the future, even if it is **current actions** that unbalance the system and eventually push it over the threshold.46 Note that these dramatic illustrations rest on assumptions that the thresholds are still relatively benign, not moving us beyond tipping points which result in an accelerated release of methane that could result in a temperature increase of more than 8 °C, possibly producing infinite impacts.47 Calculating illustrative numbers By including the welfare of future generations, something that is important when their very existence is threatened, economic discounting becomes difficult. In this chapter, some illustrative numbers are provided to indicate the order of magnitude of the values that calculations provide when traditional calculations also include future generations. These illustrative calculations are only illustrative as the timespans that must be used make all traditional assumptions questionable to say the least. Still, as an indicator for why infinite impact might be a good approximation they might help. As a species that can manipulate our environment it could be argued that the time the human race will be around, if we do not kill ourselves, can be estimated to be between 1-10 million years – the typical time period for the biological evolution of a successful species48 – and one billion years, the inhabitable time of Earth.49 [FIGURE 4 OMITTED] [FIGURE 5 OMITTED] If we assume – 50 million years for the future of humanity as our reference, – an average life expectancy of 100 years50, and – a global population of 6 billion people51 – all conservative estimate – , we have half a million generations ahead of us with a total of 3 quadrillion individuals. Assuming a value of $50,000 per life, the cost of losing them would then be $1.5 ×1020, or $150 quintillion. This is a very low estimate, and Posner suggests that maybe the cost of a life should be “written up $28 million” for catastrophic risks52. Posner’s calculations where only one future generation is included result in a cost of $336 quadrillion. If we include all future generations with the same value, $28 million, the result is a total cost of $86 sextillion, or $86 × 1021. This $86 sextillion is obviously a very rough number (using one billion years instead of 50 million would for example require us to multiply the results by 20), but again it is the magnitude that is interesting. As a reference there are about 1011 to 1012 stars in our galaxy, and perhaps something like the same number of galaxies. With this simple calculation you get 1022 to 1024, or 10 to 1,000 sextillion, stars in the universe to put the cost of infinite impacts when including future generations in perspective.53 These numbers can be multiplied many times if a more philosophical and technology-optimistic scenario is assumed for how many lives we should include in future generations. The following quote is from an article by Nick Bostrom in Global Policy Journal: “However, the relevant figure is not how many people could live on Earth but how many descendants we could have in total. One lower bound of the number of biological human life-years in the future accessible universe (based on current cosmological estimates) is 1034 years. Another estimate, which assumes that future minds will be mainly implemented in computational hardware instead of biological neuronal wetware, produces a lower bound of 1054 human-brain-emulation subjective life-years.” 54 Likewise the value of a life, $28 million, a value that is based on an assessment of how individuals chose when it comes to flying, can be seen as much too small. This value is based on how much we value our own lives on the margin, and it is reasonable to assume that the value would be higher than only a multiplication of our own value if we also considered the risk of losing our family, everyone we know, as well as everyone else on the planet. In the same way as the cost increases when a certain product is in short supply, the cost of the last humans could be assumed to be very high, if not infinite. Obviously, the very idea to put a price on the survival of humanity can be questioned for good reasons, but if we still want to use a number, $28 million per life should at least be considered as a significant underestimation. For those that are reluctant or unable to use infinity in calculations and are in need of a number for their formulas, $86 sextillion could be a good initial start for the cost of infinite impacts. But it is important to note that this number might be orders of magnitude smaller than an estimate which actually took into account a more correct estimation of the number of people that should be included in future generations as well as the price that should be assigned to the loss of the last humans. 2.3.3 Infinite impact threshold (IIT) As we address very complex systems, such as human civilisation and global ecosystems, a concept as important as infinite impact in this report is that of infinity impact threshold. This is the impact level that can trigger a chain of events that results in the end of human civilisation. The infinite impact threshold (IIT) concept represents the idea that long before an actual infinite impact is reached there is a tipping point where it (with some probability) is no longer possible to reverse events. So instead of focusing only on the ultimate impact it is important to estimate what level of impact the infinity threshold entails. The IIT is defined as an impact that can trigger a chain of events that could result first in a civilisation collapse, and then later result in an infinite impact. Such thresholds are especially important to recognise in a complex and interconnected society where resilience is decreasing. Social and ecological systems are complex, and in most complex systems there are thresholds where positive feedback loops become self-reinforcing. In a system where resilience is too low, feedback loops can result in a total system collapse. These thresholds are very difficult to estimate and in most cases it is possible only to estimate their order of magnitude. As David Orrell and Patrick McSharry wrote in A Systems Approach to Forecasting: “Complex systems have emergent properties, qualities that cannot be predicted in advance from knowledge of systems components alone”. According to complexity scientist Stephen Wolfram’s principle of computational irreducibility, the only way to predict the evolution of such a system is to run the system itself: “There is no simple set of equations that can look into its future.” 55 Orrell and McSharry also noted that “in orthodox economics, the reductionist approach means that the economy is seen as consisting of individual, independent agents who act to maximise their own utility. It assumes that prices are driven to a state of near-equilibrium by the ‘invisible hand’ of the economy. Deviations from this state are assumed to be random and independent, so the price fluctuations are often modelled using the normal distribution or other distributions with thin tails and finite variance.” The drawbacks of an approach using the normal distribution, or other distributions with thin tails and finite variance, become obvious when the unexpected happens as in the recent credit crunch, when existing models totally failed to capture the true risks of the economy. As an employee of Lehman Brothers put it on August 11, 2007: “Events that models predicted would happen only once in 10,000 years happened every day for three days.” 56 [FIGURE 6 OMITTED] The exact level for an infinite impact threshold should not be the focus, but rather the fact that such thresholds exists and that an order of magnitude should be estimated.57 During the process of writing the report, experts suggested that a relatively quick death of two billion people could be used as a tentative number until more research is available.58 With current trends undermining ecological and social resilience it should be noted that the threshold level is likely to become lower as time progress. 2.3.4 Global F-N curves and ALARP In the context of global risks with potentially infinite impact, the possibility of establishing global F-N curves is worth exploring. One of the most common and flexible frameworks used for risk criteria divides risks into three bands: 59 1. Upper: an unacceptable/ intolerable region, where risks are intolerable except in extraordinary circumstances and risk reduction measures are essential. 2. Middle: an ALARP (“as low as reasonably practicable”) region, where risk reduction measures are desirable but may not be implemented if their cost is disproportionate to the benefit achieved. 3. Lower: a broadly acceptable/ negligible region, where no further risk reduction measures are needed. The bands are expressed by F-N curves. When the frequency of events which cause at least N fatalities is plotted against the number N on log–log scales, the result is called an F-N curve.60 If the frequency scale is replaced by annual probability, then the resultant curve is called an f-N curve. The concept for the middle band when using F-N curves is ALARP. It is a term often used in the area of safety-critical and safety-involved systems.62 The ALARP principle is that the residual risk should be as low as reasonably practicable. The upper band, the unacceptable/ intolerable region, is usually the area above the ALARP area (see figure 8) By using F-N curves it is also possible to establish absolute impact levels that are never acceptable, regardless of probability (Figure 7. Based on an actual F-n Curve showing an absolute impact level that is defined as unacceptable). This has been done in some cases for local projects. The infinite threshold could be used to create an impact limit on global F-N curves used for global challenges in the future. Such an approach would help governments, companies and researchers when they develop new technical solutions and when investing in resilience. Instead of reducing risk, such an approach encourages the building of systems which cannot have negative impacts above a certain level. Pros – Clearly shows relationship between frequency and size of accident – Allows judgement on relative importance of different sizes of accident – Slope steeper than -1 provides explicit consideration of multiple fatality aversion and favours concepts with lower potential for large fatality events – Allows company to manage overall risk exposure from portfolio of all existing and future facilities Cons – Cumulative expression makes it difficult to interpret, especially by non-risk specialists – Can be awkard to derive – May be difficult to use if criterion is exceeded in one area but otherwise is well below – Much debate about criterion lines Figure 7: Example of F-n curve showing different levels of risk 61 Figure 9: Pros and cons of F-N curves 63 46 Global Challenges – Twelve risks that threaten human civilisation – The case for a new category of risks 2.3 Global challenges and infinite impact practical guidance that can provide defined group of risks 2.3.5 A name for a clearly 10 100 1000 10000 10 10 10 10 10 10 10 10-2 -3 -4 -5 -6 -7 -8 -9 Number of Fatalities (N) Frequency (F) of Accidents with N or More Fatalities (Per Year) ALARP region Unacceptable Acceptable Today no established methodology exists that provides a constantly updated list of risks that threaten human civilisation, or even all human life. Given that such a category can help society to better understand and act to avoid such risks, and better understand the relation between these risks, it can be argued that a name for this category would be helpful.65 To name something that refers to the end of humanity is in itself a challenge, as the very idea is so far from our usual references and to many the intuitive feeling will be to dismiss any such thing. The concept used in this report is “infinity”. The reson for this is that many of the challenges relate to discussed. In one way the name is not very important so long as people understand the impacts and risks associated with it. Still, a name is symbolic and can either help or make it more difficult to get support to establish the new category. The work to establish a list of risks with infinite impact evolved from “existential risk”, the philosophical concept that inspired much of the work to establish a clearly defined group of risks. The reason for not using the concept “existential risk and impact” for this category, beside the fact that existential impact is also used in academic contexts to refer to a personal impact, is that the infinite category is a smaller subset of “existential risk” and this new category is meant to be used as a tool, not a scientific concept. Not only should the impacts in the category potentially result in the end of all human life, it should be possible to affect the probability and/or impact of that risk. There must also exist an agreed methodology, such as the one suggested in this report, that decides what risks belong and not belong on the list. Another concept that the category relates to is “global catastrophic risk” as it is one of the most used concepts among academics interested in infinite impacts. However it is vague enough to be used to refer to impacts from a few thousand deaths to the end of human civilisation. Already in use but not clearly defined, it includes both the academic concept existential risks and the category of risks with infinite impacts. macroeconomics and its challenges in relation to the kind of impacts that the risks in this report focus on. Further, the name clearly highlights the unique nature without any normative judgements. Still, infinity is an abstract concept and it might not be best communicate the unique group of risks that it covers to all stakeholders. In the same way as it can be hard to use singularity to describe a black hole, it can be difficult to use infinity to describe a certain risk. If people can accept that it is only from a specific perspective that the infinity concept is relevant it could be used beyond the areas of macroeconomics. Two other concepts that also have been considered during the process of writing this report are “xrisks” and “human risk of ruin”. Xrisk has the advantage, and disadvantage, of not really saying anything at all about the risk. The positive aspect is that the name can be associated with the general concept of extinction and the philosophical concept of existential risk as both have the letter x in them. The disadvantage is the x often represents the unknown and can therefore relate to any risk. There is nothing in the name that directly relates to the kind of impacts that the category covers, so it is easy to interpret the term as just unknown risks. Human risk of ruin has the advantage of having a direct link to a concept, risk of ruin, that relates to a very specific state where all is lost. Risk of ruin is a concept in use in gambling, insurance, and finance that can all give very important contributions to the work with this new category of risk. The resemblance to an existing concept that is well established could be both a strength and a liability. Below is an overview of the process when different names were Figure 8: Example of F-n curve showing an absolute impact level that is defined as unacceptable/ infinite. i.e no level of probability is acceptable above a certain level of impact, in this case 1000 dead 64 Global Challenges – Twelve risks that threaten human civilisation – The case for a new category of risks 47 2.3 Global challenges and infinite impact 3. 2. 1. 9. Unacceptable risks in different combinations, e.g. unacceptable global risks – This is probably not appropriate for two main reasons. First, it is a normative statement and the category aims to be scientific; whether these risks are unacceptable or not is up to the citizens of the world to decide. Second, the idea of risk is that it is a combination of probability times impact. If a risk is unacceptable is therefore also usually related to how easy it is to avoid. Even if a risk is small, due to relatively low probability and relatively low impact, but is very easy to address, it can be seen as unacceptable, in the same way a large risk can be seen as acceptable if it would require significant resources to reduce. There will not be a perfect concept and the question is what concept can find the best balance between being easy to understand, acceptable where policy decisions needs to be made and also acceptable for all key groups that are relevant for work in these area. During the process to find a name for this category inspiration has been found in the process when new concepts have been introduced; from irrational numbers and genocide to sustainable development and the Human Development Index. So far “infinite risk” can be seen as the least bad concept in some areas and “xrisks” and “human risk of ruin” the least bad in others. The purpose of this report is to establish a methodology to identify a very specific group of risks as well as continue to a process where these risks will be addressed in a systematic and appropriate way. The issue of naming this group of risks will be left to others. The important is that the category gets the attention it deserves. The three concepts are very different. Global catastrophic risk is possibly the most used concept in contexts where infinite impacts are included, but it is without any clear definition. Existential risk is an academic concept used by a much smaller group and with particular focus on future technologies. The category in this report is a tool to help decision makers develop strategies that help reduce the probability that humanity will end when it can be avoided. The relation between the three concepts can be illustrated with three circles. The large circle (1) represents global catastrophic risks, the middle one (2) existential risks and the small circle (3) the list of twelve risks in this report, i.e. risks where there are peer reviewed academic studies that estimate the probability of an infinite impact and where there are known ways to reduce the risk. A list that could be called infinite risks, xrisks, or human risk of ruin. Other concepts that are related to infinite impacts that could potentially be used to describe the same category if the above suggestions are not seen as acceptable concepts are presented below, together with the main reason why these concepts were not chosen for this report. 1. Risk of ruin – is a concept in gambling, insurance and finance relating to the likelihood of losing all one’s capital or affecting one’s bankroll beyond the point of recovery. It is used to describe individual companies rather than systems.66 2. Extinction risk – is used in biology for any species that is threatened. The concept is also used in memory/cognition research. It is a very dramatic term, to be used with care. These factors make it probably unsuitable for use by stakeholders accustomed to traditional risk assessment. 3. Astronomical risk – is seldom used scientifically, but when it is used it is often used for asteroids and is probably best reserved for them.67 4. Apocalyptic risk – could have been suitable, as the original meaning is apocálypsis, from the Greek ἀπό and καλύπτω meaning ‘un-covering’. It is sometime used, but in a more general sense, to mean significant risks.68 But through history and today it is mainly used for a religious end of time scenario. Its strong links to unscientific doom-mongers make it probably unsuitable for a scientific concept. 5. End-of-the-world risk - belongs to the irrational doomsday narratives and so is probably unsuitable for scientific risk assessments. 6. Extreme risk – is vague enough to describe anything beyond the normal, so it is probably unsuitable for risk assessments of this magnitude. 7. Unique risk – is even vaguer, as every risk is unique in some way. Probably best avoided in risk assessments. 8. Collapse risk – is based on Jared Diamond’s thinking.69 There are many different kinds of collapse and only a few result in infinite impact. 48 Global Challenges – Twelve risks that threaten human civilisation – The case for a new category of risks 2.3 Global challenges and infinite impact Estimations of impact Only literature where there is some estimation of impact that indicates the possibility of an infinite impact is included. Leading organisations’ priorities In order to increase the probability of covering all relevant risks an overview of leading organisations' work was conducted. This list was then compared with the initial list and subjected to the same filter regarding the possibility to affect the probability or impact. Possibility of addressing the risk Possibility of addressing the risk: From the risks gathered from literature and organisations, only those where the probability or impact can be affected by human actions are included. Expert review Qualitative assessment: Expert review in order to increase the probability of covering all relevant global risks. List of risks Result: List of risks with potentially infinite impacts. Relevant literature Identification of credible sources: search relevant literature in academic literature included in World of Knowledge and Google Scholar. 1 2 3 4 5 6 This chapter presents the methodology used to identify global risks with potentially infinite impact. Methodology overview In order to establish a list of global risks with potentially infinite impact a methodological triangulation was used, consisting of: – A quantitative assessment of relevant literature. – A strategic selection of relevant organisations and their priorities. – A qualitative assessment with the help of expert workshops. 2.4 Methodology 70 Global Challenges – Twelve risks that threaten human civilisation – The case for a new category of risks 49 2.4 Methodology The scientific review of literature was led by Seth Baum, Executive Director of the Global Catastrophic Risk Institute72 and research scientist at the Center for Research on Environmental Decisions, Columbia University.73 The methodology for including global risks with a potentially infinite impact is based on a scientific review of key literature, with focus on peer-reviewed academic journals, using keyword search of both World of Knowledge74 and Google Scholar75 combined with existing literature overviews in the area of global challenges. This also included a snowball methodology where references in the leading studies and books were used to identify other scientific studies and books. In order to select words for a literature search to identify infinite impacts, a process was established to identify words in the scientific literature connected to global challenges with potentially infinite impacts. Some words generate a lot of misses, i.e. publications that use the term but are not the focus of this report. For example “existential risk” is used in business; “human extinction” is used in memory/cognition. Some search terms produced relatively few hits. For example “global catastrophic risk” is not used much. Other words are only used by people within a specific research community: few use “existential risk” in our sense unless they are using Nick Bostrom’s work. The term “global catastrophe” was identified as a phrase that referred almost exclusively to extremely negative impacts on humans, by a diversity of researchers, not just people in one research community. A list of 178 relevant books and reports was established based on what other studies have referred to, and/or which are seen as landmark studies by groups interviewed during the process. They were selected for a closer examination regarding the challenges they include.76 The full bibliography, even with its focus on publications of general interest, is still rather long. So it is helpful to have a shorter list focused on the highlights; the most important publications based on how often they are quoted, how wellspread the content (methodology, lists, etc.) is and how often key organisations use them. The publications included must meet at least one of the following criteria: – Historical significance. This includes being the first publication to introduce certain key concepts, or other early discussions of global challenges. Publications of historical significance are important for showing the intellectual history of global challenges. Understanding how the state of the art research got to where it is today can also help us understand where it might go in the future. – Influential in developing the field. This includes publications that are highly cited77 and those that have motivated significant additional research. They are not necessarily the first publications to introduce the concepts they discuss, but for whatever reason they will have proved important in advancing research. – State of the art. This includes publications developing new concepts at the forefront of global challenges research as well as those providing the best discussions of important established concepts. Reading these publications would bring a researcher up to speed with current research on global challenges. So they are important for the quality of their ideas. – Covers multiple global challenges (at least two). Publications that discuss a variety of global challenges are of particular importance because they aid in identifying and comparing the various challenges. This process is essential for research on global risks to identify boundaries and research priorities. In order to identify which global challenges are most commonly discussed, key surveys were identified and coded. First, a list of publications that survey at least three global challenges was compiled, and they were then scanned to find which challenges they discussed. The publications that survey many global challenges were identified from the full bibliography. Publications from both the academic and popular literature were considered. Emphasis was placed on publications of repute or other significance.78 To qualify as a survey of global challenges, the publication had to provide an explicit list of challenges or to be of sufficient length and breadth for it to discuss a variety of challenges. Many of the publications are books or book-length collections of articles published in book form or as special issues of scholarly journals. Some individual articles were also included because they discussed a significant breadth of challenges. A total of 40 global challenge survey publications were identified. For authors with multiple entries (Bostrom with three and WEF with ten) each challenge was counted only once to avoid bias. review of key literature 71 2.4.1 A scientific 50 Global Challenges – Twelve risks that threaten human civilisation – The case for a new category of risks 2.4 Methodology 0 5 10 15 20 25 Climate Change Nuclear War Pandemic Biodiversity loss Asteroid / Comet / Meteor Volcano Genetic Engineering High Energy Physics Nanotech Resource Depletion Artificial Intelligence Chemical Pollution Ecological Catastrophe Biogeochem Government Failure Poverty System Failure Astronomic Explosion LULCC Biological Weapons Chemical Weapons Extraterrestrial Reject Procreation Computer Failure EM Pulse New Technology Ozone Depletion Dysgenics Ocean Acidification Interstellar Cloud Atmosphere Aerosols Phase Transition Simulation Unknown 21 18 17 15 14 14 13 13 13 13 11 11 11 8 8 8 8 7 7 5 5 5 5 4 4 4 4 3 3 2 1 1 1 1 In terms of authorship and audience, there are 17 academic publications, 9 popular publications, 1 government report, 3 publications written by academics for popular audiences. In terms of format, there are 15 books, 5 edited collections, 7 articles, 3 of miscellaneous format. Of the 40 publications identified, 22 were available at the time of coding. In addition, 10 Global Risks Reports from the World Economic Forum were coded and then gathered under one heading: “WEF Global Risk Report 2005-2014”. A list of 34 global challenges was developed based on the challenges mentioned in the publications. A spreadsheet containing the challenges and the publications was created to record mentions of specific challenges in each publication to be coded. Then each publication was scanned in its entirety for mentions of global challenges. Scanning by this method was necessary because many of the publications did not contain explicit lists of global challenges, and the ones that did often mentioned additional challenges separately from their lists. So it was not required that a global challenge be mentioned in a list for it to be counted – it only had to be mentioned somewhere in the publication as a challenge. Assessing whether a particular portion of text counts as a global challenge and which category it fits in sometimes requires some interpretation. This is inevitable for most types of textual analysis, or, more generally, for the coding of qualitative data. The need for interpretation in this coding was heightened by the fact that the publications often were not written with the purpose of surveying the breadth of global challenges, and even the publications that were intended as surveys did not use consistent definitions of global challenges. The coding presented here erred on the side of greater inclusivity: if a portion of text was in the vicinity of a global challenge, then it was coded as one. For example, some publications discussed risks associated with nuclear weapons in a general sense without specifically mentioning the possibility of large-scale nuclear war. These discussions were coded as mentions of nuclear war, even though they could also refer to single usages of nuclear weapons that would not rate as a global challenge. This more inclusive approach is warranted because many of the publications were not focused exclusively on global challenges. If they were focused on them, it is likely that they would have included these risks in their global challenge form (e.g., nuclear war), given that they were already discussing something related (e.g., nuclear weapons). Below are the results from the overview of the surveys. Figure 9: Number of times global challenges are included in surveys of global challenges Global Challenges – Twelve risks that threaten human civilisation – The case for a new category of risks 51 2.4 Methodology Climate Change Nuclear War Pandemic Biodiversity loss Asteroid / Comet / Meteor Volcano Genetic Engineering High Energy Physics Nanotech Resource Depletion Artificial Intelligence Chemical Pollution Ecological Catastrophe 21 18 17 15 14 14 13 13 13 13 11 11 11 0 25 20 15 10 5 dung beetle star trek zinc oxalate human extinction 0 200 400 600 800 1000 It should be noted that the literature that includes multiple global challenges with potentially infinite impact is very small, given the fact that it is about the survival of the human race. Experts in the field of global challenges, like Nick Bostrom, have urged policymakers and donors to focus more on the global challenges with infinite impacts and have used dramatic rhetoric to illustrate how little research is being done on them compared with other areas. However, it is important to note that many more studies exist that focus on individual global risks, but often without including low-probability high-impact outcomes.80 How much work actually exists on human extinction infinite impact is therefore difficult to assess. The list of risks found in the scientific literature was checked against a review of what challenges key organisations working on global challenges include in their material and on their webpages. This was done to ensure that no important risk was excluded from the list. The coding of key organisations paralleled the coding of key survey publications. Organisations were identified via the global catastrophic risk organisation directory published by the Global Catastrophic Risk Institute.82 They were selected from the directory if they worked on a variety of global challenges – at least three, and ideally more. The reason for focusing on those that work on multiple challenges is to understand which challenges they consider important and why. In contrast, organisations that focus on only one or two challenges may not Figure 10: The global challenges included ten times or more in surveys of global challenges on global challenges 81 organisations working 2.4.2 A review of Figure 11: Number of academic papers on various topics (listed in Scopus, August 2012) From the paper “Existential Risk Prevention as Global Priority” 79 52 Global Challenges – Twelve risks that threaten human civilisation – The case for a new category of risks 2.4 Methodology Climate Change Nuclear War Pandemic Resource Depletion Biological Weapons Computer Failure Government Failure Nanotech Chemical Weapons Artificial Intelligence Genetic Engineering System Failure Biodiversity loss Ecological Failure Poverty Volcano Asteroid / Comet / Meteor Astronomic Explosion Biogeochem Chemical Pollution Extraterrestrial High Energy Physics New Technology Ozone Depletion Atmospheric Aerosols Dysgenics EM Pulse Interstellar Cloud LULCC Ocean Acidification Phase Transition Reject Procreation Simulation Unknown 13 13 12 9 8 7 7 7 6 5 4 4 2 2 2 2 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 4 8 12 2 6 10 14 be able to adjust their focus according to which challenges they consider the most important. The organisation coding used the same coding scheme developed for coding survey publications. References to specific global challenges were obtained from organisations’ websites. Many have web pages which list the topics they work on. Where possible, references to global challenges were pulled from these pages. Additional references to these challenges were identified by browsing other web pages, including recent publications. While it is possible that some of these organisations have worked on global challenges not mentioned on the web pages that were examined, overall the main challenges that they have worked on have probably been identified and coded. So the results should give a reasonably accurate picture of what global challenges these organisations are working on. Organisations working with global challenges were initially selected on the basis of the literature overview. A snowball sampling was conducted based on the list of organisations identified, according to whether they claimed to work on global challenges and/or their web page contained information about “existential risk”, “global catastrophic risk”,“human extinction” or “greatest global challenges”. Cross-references between organisations and input during the workshops were also used to identify organisations. An initial list of 180 organisations which work with global challenges was established. Based on the production of relevant literature, which other organisations referred to the organisation, and/or are seen as influential by groups interviewed during the process, a short-list of organisations were selected for a closer examination regarding the challenges they work with. Then those working with multiple challenges were selected, resulting in a list of 19 organisations.83 Below is the overview of the results from the overview of key organisations working with multiple global challenges. The organisations working on global challenges vary widely in: 1. What they count as a global challenge 2. How systematically they identify global challenges; and 3. Their emphasis on the most important global challenges For most organisations working with global challenges there are no explanations for the methodology used to select the challenges. Only a few thought leaders, like Tower Watson and their Extreme Risk Report 2013, have a framework for the challenges and estimates of possible impacts. Figure 12: Global challenges that key organisations work with Global Challenges – Twelve risks that threaten human civilisation – The case for a new category of risks 53 2.4 Methodology Climate Change Nuclear War Pandemic Resource Depletion Biological Weapons Computer Failure Government Failure Nanotech Chemical Weapons Artificial Intelligence Genetic Engeneering System Failure Atmospheric Aerosols 13 13 12 9 8 7 7 7 6 5 4 4 0 4 8 12 2 6 10 14 In most cases there is neither a definition of the impact, nor a definition of the probability. The report that focuses on global risk which is probably best known is the WEF Global Risk Report. The WEF’s risk work, with many other groups’, is probably best described as belonging to the category of risk perception rather than risk assessment, where experts are asked to estimate risks, but without any clear definition of probability or impact. The more serious organisations, like the WEF, also clearly define what they do as discussing perception of risk, not a scientific assessment of the actual risk. The WEF describes its perception methodology as follows: “This approach can highlight areas that are of most concern to different stakeholders, and potentially galvanise shared efforts to address them.” 85 The question which people are asked to answer is: “What occurrence causes significant negative impact for several countries and industries?” 86 The respondents are then asked to provide a number on two scales from 1-4, one for impact and another for likelihood (within 10 years).87 It is then up to the respondent to define what 1-4 means, so the major value of the report is to track the changes in perception over the years. Such perception approaches are obviously very interesting and, as the WEF states, can influence actual probability as the readers’ decisions will be influenced by how different challenges are perceived. Still, it is important to remember that the report does not provide an assessment of the actual probability (0-100%) or an assessment of the impact (and not the impact on human suffering, as many respondents likely define risk in monetary terms for their own company or country). An overview of WEF reports from the last ten years indicates that the challenges that likely could happen when applying a five year horizon, like the first signs of climate change, governmental failure and traditional pandemic, are identified. On the other hand, challenges which have very big impacts but lower probability, like extreme climate change, nanotechnology, major volcanoes, AI, and asteroids, tend to get less, or no, attention. An important question to explore is whether a focus on the smaller but still serious impacts of global challenges can result in an increased probability of infinite impacts. For example, there are reasons to believe that a focus on incremental adaptation instead of significant mitigation could be a problem for climate change as it could result in high-carbon lock-in.88 Other research indicates that focus on commercially relevant smaller pandemics could result in actions that make a major pandemic more likely. It is argued that this could happen, for example, by encouraging increased trade of goods while investing in equipment that scans for the type of pandemics that are known. Such a system can reduce the probability for known pandemics while at the same time resulting in an increased probability for new and more serious pandemics.89 Figure 13: The top 12 global challenges that key organisations work with 2.4.3 Workshops global risks 2.5 The list of Two workshops were arranged where the selection of challenges was discussed, one with risk experts in Oxford at the Future of Humanity Institute and the other in London with experts from the financial sector. See Appendix 2 for agenda and participants. In both workshops the list of global challenges was discussed to see if any additional challenges should be included, or if there were reasons to exclude some from the list. No challenge was excluded at the workshops, but one was added. Although little research exists yet that is able to verify the potential impacts, the participants agreed to include Global System Collapse as a risk with possible infinite impact. There was agreement that further research is needed to clarify exactly what parts of the economic and political system could collapse and result in a potentially infinite outcome. The conclusion was that enough research exists to include such a collapse on the list. Based on the risks identified in the literature review and in the review of organisations and applying the criteria for potentially infinite impact, these risks were identified: 1. Extreme Climate Change 2. Nuclear War 3. Global Pandemic 4. Ecological Catastrophe 5. Global System Collapse 6. Major Asteroid Impact 7. Supervolcano 8. Synthetic Biology 9. Nanotechnology 10. Artificial Intelligence (AI) 11. Unknown Consequences 12. Future Bad Global Governance This is an initial list. Additional risks will be added as new scientific studies become available, and some will be removed if steps are taken to reduce their probability90 and/or impact so that they no longer meet the criteria. Four categories of global challenges The challenges included in this report belong to four categories. The first, current challenges, includes those where decisions today can result directly in infinite impacts. They are included even if the time between action and impact might be decades, as with climate change. The second category is exogenous challenges, those where decisions do not – currently – influence probability, but can influence impact. The third category is emerging challenges, those where technology and science are not advanced enough to pose a severe threat today, but where the challenges will probably soon be able to have an infinite impact. The technologies included in emerging challenges, including synthetic biology, nanotechnology and artificial intelligence (AI), will be critical to finding solutions to infinite impacts. Including these technologies should not be seen as an attempt to arrest them. If anything, the development of sustainable solutions should be accelerated. But it is equally important to create guidelines and frameworks to avoid their misuse, whether intentional or accidental. The fourth category, future global policy challenges, is of a different kind. It includes challenges related to the consequences of an inferior or destructive global governance system. This is especially important as well-intended actions to reduce global challenges could lead to future global governance systems with destructive impact. The first category, current challenges, includes: 1. Extreme Climate Change 2. Nuclear War 3. Global Pandemic 4. Ecological Catastrophe 5. Global System Collapse The second category, exogenous challenges, covers: 6. Major Asteroid Impact 7. Supervolcano Those in the third category, emerging challenges, are: 8. Synthetic Biology 9. Nanotechnology 10. Artificial Intelligence (AI) 11. Unknown Consequences The fourth category, global policy challenges, is: 12. Future Bad Global Governance not included 2.5.1 Risks Many risks could severely damage humanity but have not been included in this report. They were excluded for one or more of three reasons: 1. Limited impact. Many challenges can have significant local negative effects, without approaching the “2 billion negatively affected” criterion - tsunamis, for example, and chemical pollution. 2. No effective countermeasures. The report focuses on promoting effective interventions and so ignores challenges where nothing useful can be done to prevent or mitigate the impact, as with nearby gamma-ray bursts. 3. Included in other challenges. Many challenges are already covered by others, or have a damage profile so similar that there seemed no need to have a separate category. Population growth, for one, is an underlying driver significant for climate change and eco-system catastrophe, but without direct large-scale impacts. The challenges mentioned in the reviewed literature and organisations which are not included in this report often refer to economic damage such as “fiscal crises” or “unemployment”. While such impacts could have far-reaching consequences they are obviously of another magnitude than those included here. Some of the risks that were suggested and/or which exist in books and reports about global risks were rejected according to the criteria above. They include: 91 1. Astronomical explosion/nearby gamma-ray burst or supernova.92 These seem to be events of extremely low probability and which are unlikely to be survivable. Milder versions of them (where the source is sufficiently far away) may be considered in a subsequent report. ͢ Not included due to: No effective countermeasures 2. False vacuum collapse. If our universe is in a false vacuum and it collapses at any point, the collapse would expand at the speed of light destroying all organised structures in the universe.93 This would not be survivable. ͢ Not included due to: No effective countermeasures 3. Chemical pollution. Increasingly, there is particular concern about three types of chemicals: those that persist in the environment and accumulate in the bodies of wildlife and people, endocrine disruptors that can interfere with hormones, and chemicals that cause cancer or damage DNA. ͢ Not included due to: Limited impact 4. Dangerous physics experiments creating black holes/strangelets including high energy physics. These risks are of low probability94 and have been subsumed under “Uncertain Risks”. ͢ Not included due to: Included in other challenges 5. Destructive solar flares. Though solar flares or coronal mass ejections could cause great economic damage to our technological civilisation,95 they would not lead directly to mass casualties unless the system lacks basic resilience. They have been subsumed in the Global System Collapse category. ͢ Not included due to: Limited impact/included in other challenges 6. Moral collapse of humanity. Humanity may develop along a path that we would currently find **morally repellent**. The consequences of this are **not clear-cut**, and **depend on value judgements that would be contentious and unshared**.96 Some of these risks (such as global totalitarianism or **enduring poverty**) were included in the Governance Disasters category. ͢ Not included due to: included in other challenges 7. Resource depletion/LULCC/ Biodiversity loss. It has often been argued that declining resources will cause increased conflict.97 Nevertheless such conflicts would **not** be **sufficient in themselves to threaten humanity on a large scale, without a “ System Collapse”** or “Governance Disasters”. ͢ Not included due to: included in other challenge

**Existential risk first – prefer utilitarianism**

**Bostrom 12** (Nick, Professor of Philosophy at Oxford, directs Oxford's Future of Humanity Institute and winner of the Gannon Award, Interview with Ross Andersen, correspondent at The Atlantic, 3/6, “We're Underestimating the Risk of Human Extinction”, <http://www.theatlantic.com/technology/archive/2012/03/were-underestimating-the-risk-of-human-extinction/253821/>)

Bostrom, who directs Oxford's Future of Humanity Institute, has argued over the course of several papers that human extinction risks are poorly understood and, worse still, **severely underestimated by society**. Some of these existential risks are fairly well known, especially the natural ones. But others are obscure or even exotic. Most worrying to Bostrom is the subset of existential risks that arise from human technology, a subset that he expects to grow in number and potency over the next century.¶ Despite his concerns about the risks posed to humans by technological progress, Bostrom is no luddite. In fact, he is a longtime advocate of transhumanism---the effort to improve the human condition, and even human nature itself, through technological means. In the long run he sees technology as a bridge, a bridge we humans must cross with great care, in order to reach new and better modes of being. In his work, Bostrom uses the tools of philosophy and mathematics, in particular probability theory, to try and determine how we as a species might achieve this safe passage. What follows is my conversation with Bostrom about some of the most interesting and worrying existential risks that humanity might encounter in the decades and centuries to come, and about what we can do to make sure we outlast them.¶ Some have argued that we ought to be directing our resources toward humanity's existing problems, rather than future existential risks, because many of the latter are highly improbable. You have responded by suggesting that **existential risk mitigation may in fact be a dominant moral priority over the alleviation of present suffering**. Can you explain why? ¶ Bostrom: Well suppose you have a moral view that counts future people as being worth as much as present people. You might say that fundamentally it doesn't matter whether someone exists at the current time or at some future time, just as many people think that from a fundamental moral point of view, it doesn't matter where somebody is spatially---somebody isn't automatically worth less because you move them to the moon or to Africa or something. A human life is a human life. If you have that moral point of view that future generations matter in proportion to their population numbers, then you get this **very stark implication that existential risk mitigation has a much higher utility than pretty much anything else that you could do**. There are so many people that could come into existence in the future if humanity survives this critical period of time---we might live for billions of years, our descendants might colonize billions of solar systems, and there could be billions and billions times more people than exist currently**. Therefore, even a very small reduction in the probability of realizing this enormous good will tend to outweigh even immense benefits like eliminating poverty or curing malaria**, which would be tremendous under ordinary standards.

**Framing**

**Probability First**

**Risk is multiplicative -- the conjunctive fallacy mitigates the risk of DAs.**

**Conetta 98** (Carl Conetta, Director of the Project on Defense Alternatives, Research Fellow of the Institute for Defense and Disarmament Studies, researcher and awarded author at the Pentagon, US State Department, US House Armed Services Committee, Army War College, National Defense University, and UNIDIR, March 1998, "Global Beat: Dueling with Uncertainty: The New Logic of American Military Planning," Project on Defense Alternatives, http://www.bu.edu/globalbeat/usdefense/conetta0398.html)

Playing with Wild Cards

Without doubt, simulations -- including nonstandard ones -- can aid planning. The question is: To what end? And to what effect? Exploring "wild cards" in order to identify warning signs or to define limits is one thing; using them to establish force structure or modernization requirements, quite another. Especially suspect would be using scenarios that are detached from declared US interests to define current requirements; this would put the military "cart" before the political "horse." Another, broader concern is how the effusion of improbable conflict scenarios affects public policy discourse overall.

Conflict scenarios, both wild and tame, can gain **more credibility** in the telling than they deserve. Cognitive researcher Massimo Piattelli-Palmarini calls this the "Othello effect," referring to the trail of **plausible** but **false suppositions** that led Othello to murder his wife, Desdemona. Even the most **farfetched scenarios** comprise a **number of steps or links** each of which may seem plausible or even probable given the one that came before. Although the likelihood of the scenario **dwindles with each step**, the residual impression is one of plausibility. Omitted are the many branches at **each step** that would lead to a neutral or even positive outcome. The resulting **snapshots**, although numerous, offer a **highly-selective** view of what the future may hold. And the fact that only the negative outcomes are articulated and exercised can **distort the** general public impression of risk.

Living with Uncertainty

There is no escape from uncertainty, but there is relief from uncertainty hysteria. It begins with recognizing that instability has boundaries -- just as turbulence in physical systems has discernable onset points and parameters. The turbulence of a river, for instance, corresponds to flow and to the contours of the river's bed and banks. It occurs in patches and not randomly. The weather also is a chaotic system that resists precise long-range forecasting, but allows useful prediction of broader trends and limits.

Despite uncertainty, statements of probability matter. They indicate the weight of evidence -- or whether there is any evidence at all. The uncertainty hawks would flood our concern with a horde of dangers that pass their permissive test of "**non-zero probability**." However, by **lowering the threshold** of alarm, they establish an **impossible standard of defense** sufficiency: absolute and certain military security. Given finite resources and competing ends, something less will have to do. Strategic wisdom begins with the setting of priorities -- and priorities demand strict attention to what appears likely and what does not.

**Each life improved solves an existential catastrophe -- the butterfly effect outweighs.**

**Kaczmarek 17** – Patrick Kaczmarek, PhD at the University of Glasgow, a Senior Researcher at Effective Giving, Visiting Researcher at the Future of Humanity Institute at the University of Oxford and a Visiting Scholar at the Department of Philosophy at the University of Pittsburgh. [How Much is Rule-Consequentialism Really Willing to Give Up to Save the Future of Humanity? Utilitas, 29(2), https://www.cambridge.org/core/journals/utilitas/article/how-much-is-ruleconsequentialism-really-willing-to-give-up-to-save-the-future-of-humanity/F867301151A79F7DA566A14DF71749B3]//BPS

Notice, the problem can be cast two different ways. First, the loss associated with humanity's premature extinction is so great that even if the probability of a catastrophic event is very low, an expected value calculation suggests that we should strive to prevent its possible occurrence. And yet, there is something **deeply puzzling** about **ruining** the lives of all **actual persons** for the sake of humanity eking out a **longer stay** in the universe.

Second, you may have realized that the above implication bears **close resemblance** to **the dreaded Repugnant Conclusion**. The Repugnant Conclusion states that for any population, all with a very high quality of life, there must be some larger imaginable population whose existence, all else being equal, would be better despite their lives being barely worth living.19The mistake, as countless critics have noted, is that quantity (that is, size of population) should not be able to compensate for a stark reduction to their average quality of life.

I'm inclined to agree that this looks worrisome. For some, if this were the end of the story, it would surely act as a reductio ad absurdum of the view. But this is not the full story.

**AN INDIRECT APPROACH** TO LOWERING THE **THREAT OF EXTINCTION**

In setting out our earlier comparison of the two populations it was assumed that only costs go up, never benefits. That is to say, A was fixed and the total sum of goods went up merely because the size of the population grew, despite internalization costs reducing average quality of life. Colouring in the picture, this corresponds to the scenario where, all else being equal, existential threats are directly targeted. To illustrate, this could amount to putting a lot of resources towards asteroid deflection programmes.20

I now wish to argue that we could **instead reduce** existential risk by **indirect means**, and in so doing make the world in **two ways go better**. As noted earlier, we would prolong humanity's place in the cosmos. Furthermore, an indirect approach improves the average welfare of persons, particularly **the worse-off in our population**.

Certainly, it would be a mistake to concentrate exclusively on indirectly lowering the probability of doomsday. Returning to our earlier example, reducing global poverty cannot prevent an Earth-bound asteroid the size of Texas from making impact. Nevertheless, if we were also to adopt an **indirect approach**, then this would contribute to **existential risk reduction** by curbing the **negative ripple effects** of readily preventable illnesses, global hunger, and so forth.

Ripple effects are a class of phenomena that affect the **far future** in significant ways, shaping how our history unfolds over time.21A ripple effect is initiated by a particular event that has some causal influence on the course of events that follow it. These events, in turn, may have their own impact on how further events play out. And so on it goes, reaching **wider and wider** as time passes.

Consider the following example. A doctor is in a position to cure some infant's blindness. Sure, the infant will probably have a better life after the operation. Most of us are quick to hone-in on this feature of the situation. And many other goods go unacknowledged by us as a result. Just a few of the proximate advantages we might reasonably expect to find after curing the infant's blindness include: her parents will be less worried about her, subsequently finding more free time to develop their own personal projects; the government will spend fewer resources on providing her education; this child will grow up with more opportunities, as well as perhaps being inspired to start a grassroots initiative or develop an **anti-malarial drug**. All of these consequences will have some role in shaping our future due to their own ripple effects. This **network of ripple effects** might go so far as causing '[her] country's economy to develop very slightly more quickly, or make **certain technological or cultural innovations** arrive **more quickly'**.22

**Infinite risk does not justify ignoring probability---AND collapses decision-making.**

**Kessler ‘8** (Oliver; April 2008; Ph.D. in IR, Professor of Sociology at the University of Bielefeld, Professor of History and Theory of IR at the Faculty of Arts; Alternatives, Vol. 33, “From Insecurity to Uncertainty: Risk and the Paradox of Security Politics” p. 211-232; RP)

The problem of the second method is that it is very difficult to "calculate" politically unacceptable losses. If the risk of terrorism is defined in traditional terms by probability and potential loss, then the focus on dramatic terror attacks leads to the **marginalization of probabilities**. The reason is that even the highest degree of improbability **becomes irrelevant** as the measure of loss **goes to infinity**. ^o The mathematical calculation of the risk of terrorism thus tends to overestimate and to dramatize the danger. This has consequences beyond the actual risk assessment for the formulation and execution of "risk policies": If one factor of the risk calculation approaches infinity (e.g., if a case of nuclear terrorism is envisaged), then there is no balanced measure for antiterrorist efforts, and **risk management as a rational endeavor breaks down**. Under the historical condition of bipolarity, the "ultimate" threat with nuclear weapons could be balanced by a similar counterthreat, and new equilibria could be achieved, albeit on higher levels of nuclear overkill. Under the new condition of uncertainty, no such rational balancing is possible since knowledge about actors, their motives and capabilities, is largely absent. The second form of security policy that emerges when the deterrence model collapses mirrors the "social probability" approach. It represents a **logic of catastrophe**. In contrast to risk management framed in line with logical probability theory, the logic of catastrophe does not attempt to provide means of absorbing uncertainty. Rather, it takes uncertainty as constitutive for the logic itself; uncertainty is a crucial precondition for catastrophes. In particular, catastrophes happen at once, without a warning, but with major implications for the world polity. In this category, we find the impact of meteorites. Mars attacks, the tsunami in South East Asia, and 9/11. To conceive of terrorism as catastrophe has consequences for the formulation of an adequate security policy. Since catastrophes hap-pen irrespectively of human activity or inactivity, **no political action** could possibly prevent them. Of course, there are precautions that can be taken, but the framing of terrorist attack as a catastrophe points to spatial and temporal characteristics that are beyond "rationality." Thus, political decision makers are exempted from the responsibility to provide security—as long as they at least try to preempt an attack. Interestingly enough, 9/11 was framed as catastrophe in various commissions dealing with the question of who was responsible and whether it could have been prevented. This makes clear that under the condition of uncertainty, there are no objective criteria that could serve as an anchor for measuring dangers and assessing the quality of political responses. For ex- ample, as much as one might object to certain measures by the US administration, it is almost impossible to "measure" the success of countermeasures. Of course, there might be a subjective assessment of specific shortcomings or failures, but there is no "common" currency to evaluate them. As a consequence, the framework of the security dilemma fails to capture the basic uncertainties. Pushing the door open for the security paradox, the main problem of security analysis then becomes the question how to integrate dangers in risk assessments and security policies about which simply nothing is known. In the mid 1990s, a Rand study entitled "New Challenges for Defense Planning" addressed this issue arguing that "most striking is the fact that we do not even know who or what will constitute the most serious future threat, "^i In order to cope with this challenge it would be essential, another Rand researcher wrote, to break free from the "tyranny" of plausible scenario planning. The decisive step would be to create "discontinuous scenarios ... in which there is **no plausible** audit **trail** or storyline from current events"52 These nonstandard scenarios were later called "wild cards" and became important in the current US strategic discourse. They justified the transformation from a threat-based toward a capability- based defense planning strategy.53 The problem with this kind of risk assessment is, however, that even the most **absurd scenarios** can gain plausibility. By constructing a **chain of potentialities**, improbable events are linked and brought into the realm of the possible, if not even the probable. "Although the likelihood of the scenario dwindles with each step, the residual impression is one of plausibility. "54 This so-called Othello effect has been effective in the dawn of the recent war in Iraq. The connection between Saddam Hussein and Al Qaeda that the US government tried to prove was disputed from the very beginning. False evidence was again and again presented and refuted, but this did not prevent the administration from presenting as the main rationale for war the improbable yet possible connection between Iraq and the terrorist network and the improbable yet possible proliferation of an improbable yet possible nuclear weapon into the hands of Bin Laden. As Donald Rumsfeld famously said: "**Absence of evidence is not evidence of absence**." This sentence indicates that under the condition of genuine uncertainty, different evidence criteria prevail than in situations where security problems can be assessed with relative certainty.

**Prioritizing magnitude collapses into irrationality and inaction.**

**Munthe 15** – Christian Munthe, PhD, Practical Philosophy Professor Associate Head of Department for Research at the University of Gothenburg. [Why Aren't Existential Risk / Ultimate Harm Argument Advocates All Attending Mass? Philosophical Comment, 2-1-15, http://philosophicalcomment.blogspot.com/2015/02/why-arent-existential-risk-ultimate.html]//BPS

An increasingly popular genre in the sort of applied philosophy and ethics of technology, which does not so much engage with actual technological development as more or less wild phantasies about possibly forthcoming ones is the notions of "existential risks" or "ultimate harms", or similar expressions. The theme is currently inspiring several research environments at world-leading universities, such as this one and this one (where you can find many links to other sources, articles, blog posts, and so on), and given quite a bit of space in recent scholarly literature on a topic often referred to as the ethics of emerging technology. Now, personally and academically, as it has actually proceeded, I have found much of this development being to a large extent a case of the emperor's new clothes. The fact that there are possible threats to human civilizations, the existence of humanity, life on earth or, at least, extended human well-being, is not exactly news, is it? Neither is there any kind of new insight that some of these are created by humans themselves. Also, it is not any sort of recent revelation that established moral ideas, or theories of rational decision making, may provide reason for avoiding or mitigating such threats. Rather, both these theses follow rather trivially from a great many well-established ethical and philosophical theories, and are well-known to do so since hundreds of years. Still, piece after piece is being produced in the existential risk genre making this out as some sort of recent finding, and exposing grand gestures at proving the point against more or less clearly defined straw-men.

At the same time, quite a bit of what is currently written on the topic strikes me as **philosophically shallow**. For instance, the notion that the eradication of the human species has to be a bad thing seems to be far from obvious from a philosophical point of view - this would depend on such things as the source of the value of specifically human existence, the manner of the imagined extinction (it certainly does not have to involve any sort of carnage or catastrophe), and what might possibly come instead of humanity or currently known life when extinct and how that is to be valued. Similarly, it is a very common step in the typical existential risk line to jump rather immediately from the proposition of such a risk to the suggestion that substantial (indeed, massive) resources should be spent on its prevention, mitigation or management. This goes for everything from imagined large scale geo-engineering solutions to environmental problems, dreams of outer space migration, to so-called human enhancement to adapt people to be able to handle otherwise massive threats in a better way. At the same time, the advocates of the existential risk line of thought also urges caution in the application of new hitherto unexplored technology, such as synthetic biology or (if it ever comes to appear) "real" A.I. and android technology. However, also there, the angle of analysis is often restricted to this very call, typically ignoring the already since long ongoing debates in the ethics of technology, bioethics, environmental ethics, et cetera, where the issue of how much of and what sort of such caution may be warranted in light of various good aspects of different the technologies considered. And, to be frank, this simplification seems to be the only thing that is special with the existential risk argument advocacy: the idea that the mere possibility of a catastrophic scenario justifies substantial sacrifices, without having to complicate things by pondering alternative uses of resources.

Now, this kind of argument, is (or should be) **well-known** to anyone with a philosophical education, since it seems to share the basic form of the **philosophical classic known as Pascal's Wager**. In this argument, French enlightenment philosopher and mathematician, Blaise Pascal offered a "proof" of the rationality of believing in God (the sort of God found in abrahamitic monotheistic religion, that is), based on the possible consequences of belief or non-belief, given the truth or falsity of the belief. You can explore the details of Pascal's argument, but the basic idea is that in the face of the immense consequences of belief and non-belief if God exists (eternal salvation vs. eternal damnation), it is rational to bet on the existence of God, no matter what theoretical or other evidence for the truth of this belief exists and no matter the probability of this truth. It seems to me that the **typical existential risk argument advocacy** subscribes to a **very similar logic**. For instance, the standard line to defend that resources should be spent on probing and (maybe) facilitating), e.g., possible extraterrestial migration for humanity, seems to have the following form:

1) Technology T might possibly prevent/mitigate existential risk, E

2) It would be really, really, very, very bad if E was to be actualised

3) Therefore: If E was otherwise to be actualised, it would be really, really, very, very good if E was prevented

4) Therefore: If E was otherwise to be actualised, it would be really, really, very, very good if we had access to a workable T

5) Therefore: there are good reasons to spend substantial resources on probing and (maybe, if that turns out to be possible) facilitating a workable T

That is, what drives the argument is the (mere) possibility of a massively significant outcome, and the (mere) possibility of a way to prevent that particular outcome, thus doing masses of good. Now, I'm sure that everyone can see that this argument is far from obviously valid, even if we ignore the question of whether or not premise 2 is true, and this goes for Pascal's Wager too in parallel ways. For instance, the existential risk argument above seems to **ignore** that there seems to be an **innumerable amount** of thus **(merely) possible existential risk scenarios**, as well as **innumerable (merely) possibly workable technologies** that might help to prevent or mitigate each of these, and it is **unlikely (to say the least)** that we have resources to **bet substantially** on them **all**, unless we spread them **so thin** that **this action** becomes **meaningless**. Similarly, there are innumerable possible versions of the god that lures you with threats and promises of damnation and salvation, and what that particular god may demand in return, often implying a ban on meeting a competing deity's demands, so the wager doesn't seem to tell you to try to start believing in any particular of all these (merely) possible gods. Likewise, the argument above **ignores completely** the **(rather high) likelihood** that the mobilised resources will be **mostly wasted**, and that, therefore, there are **substantial opportunity costs** attached to not using these resources to use **better proven strategies** with better identified threats and problems (say, preventing global poverty) - albeit maybe not as **massive** as the outcomes in the **existential risk scenarios**. Similarly, Pascal's Wager completely ignores all the good things one needs to give up to meet the demands of the god promising eternal salvation in return (for instance, spending your Sundays working for the allieviation of global poverty). None of that is worth **any consideration**, the idea seems to be, in light of the **massive stakes** of the **existential risk** / **religious belief** or **non-belief scenarios**.

Now, I will not pick any quarrel with the existential risk argument as such on these grounds, although I do think that more developed ways to analyse risk-scenarios and the ethical implications of these already in existence and used in the fields I referred above will mean lots of troubles for the simplistic aspects already mentioned. What I do want to point to, however, is this: If you're impressed by the existential risk argument, you should be **equally impressed** by Pascal's Wager. Thus, in accordance with Pascal's recommendation that authentic religious belief can be gradually installed via the practice of rituals, you should – as should indeed the existential risk argument advocates themselves – spend your Sundays celebrating mass (or any other sort ritual demanded by the God you bet on). I **very much doubt**, however, that you (or they) in fact do that, or even accept the conclusion that you (or they) should be doing that.

**Predictions Fail**

**IR is not a social science—and even if it were, that wouldn’t be nearly enough for objective evaluation**

**BERNSTEIN ET AL. ’00** --- Steven , Richard Ned Lebow, Janice Gross Stein and Steven Weber, University of Toronto, The Ohio State University, University of Toronto and University of California at Berkeley. European Journal of International Relations 2000; 6; 43.

A deep irony is embedded in the history of the scientific study of international relations. Recent generations of scholars separated policy from theory to gain an intellectual distance from decision-making, in the belief that this would enhance the 'scientific' quality of their work. But five decades of well-funded efforts to develop theories of international relations have produced precious little in the way of useful, high confidence results. Theories abound, but few meet the most relaxed 'scientific' tests of validity. Even the most robust generalizations or laws we can state — war is more likely between neighboring states, weaker states are less likely to attack stronger states — are close to trivial, have important exceptions, and for the most part stand **outside any consistent body of theory**. A generation ago, we might have excused our performance on the grounds that we were a young science still in the process of defining problems, developing analytical tools and collecting data. This excuse is neither credible nor sufficient; there is no reason to suppose that another 50 years of well-funded research would result in anything resembling a valid theory in the Popperian sense. We suggest that the nature, goals and criteria for judging social science theory should be rethought, if theory is to be more helpful in understanding the real world. We begin by justifying our pessimism, both conceptually and empirically, and argue that the quest for predictive theory **rests on a mistaken analogy** between physical and social phenomena. Evolutionary biology is a more productive analogy for social science. We explore the value of this analogy in its 'hard' and 'soft' versions, and examine the implications of both for theory and research in international relations.' We develop the case for forward `tracking' of international relations on the basis of local and general knowledge as an alternative to backward-looking attempts to build deductive, nomothetic theory. We then apply this strategy to some emerging trends in international relations. This article is not a nihilistic diatribe against 'modern' conceptions of social science. Rather, it is a plea for constructive humility in the current context of attraction to deductive logic, falsifiable hypothesis and large- n statistical 'tests' of narrow propositions. We propose a practical alternative for social scientists to pursue in addition, and in a complementary fashion, to `scientific' theory-testing. Physical and chemical laws make two kinds of predictions. Some phenomena — the trajectories of individual planets — can be predicted with a reasonable degree of certainty. Only a few variables need to be taken into account and they can be measured with precision. Other mechanical problems, like the break of balls on a pool table, while subject to deterministic laws, are inherently unpredictable because of their complexity. Small differences in the lay of the table, the nap of the felt, the curvature of each ball and where they make contact, amplify the variance of each collision and lead to what appears as a near random distribution of balls. Most predictions in science are probabilistic, like the freezing point of liquids, the expansion rate of gases and all chemical reactions. Point predictions appear possible only because of the large numbers of units involved in interactions. In the case of nuclear decay or the expansion of gases, we are talking about trillions of atoms and molecules. In international relations, even more than in other domains of social science, it is often impossible to assign metrics to what we think are relevant variables (Coleman, 1964: especially Chapter 2). The concepts of polarity, relative power and the balance of power are among the most widely used independent variables, but there are **no commonly accepted definitions or measures** for them. Yet without consensus on definition and measurement, almost every statement or **hypothesis will have too much wiggle room to be `tested'** decisively against evidence. What we take to be dependent variables fare little better. Unresolved controversies rage over the definition and evaluation of deterrence outcomes, and about the criteria for democratic governance and their application to specific countries at different points in their history. Differences in coding for even a few cases have significant implications for tests of theories of deterrence or of the democratic peace (Lebow and Stein, 1990; Chan, 1997). The lack of consensus about terms and their measurement is not merely the result of intellectual anarchy or sloppiness — although the latter cannot entirely be dismissed. Fundamentally, it has more to do with the arbitrary nature of the concepts themselves. Key terms in physics, like mass, temperature and velocity, refer to aspects of the physical universe that we cannot directly observe. However, they are embedded in theories with deductive implications that have been verified through empirical research. Propositions containing these terms are legitimate assertions about reality because their truth-value can be assessed. Social science theories are for the most part **built on 'idealizations', that is, on concepts that cannot be anchored to observable phenomena** through rules of correspondence. Most of these terms(e.g. rational actor, balance of power) are not descriptions of reality but implicit 'theories' about actors and contexts that do not exist (Hempel, 1952; Rudner, 1966; Gunnell, 1975; Moe, 1979; Searle, 1995: 68-72).The inevitable differences in interpretation of these concepts lead to different predictions in some contexts, and these outcomes may eventually produce widely varying futures (Taylor, 1985: 55). If problems of definition, measurement and coding could be resolved, we would still find it difficult, if not impossible, to construct large enough samples of comparable cases to permit statistical analysis. It is now almost generally accepted that in the analysis of the causes of wars, the variation across time and the complexity of the interaction among putative causes make the likelihood of a general theory extraordinarily low. Multivariate theories run into the problem of negative degrees of freedom, yet international relations rarely generates data sets in the high double digits. Where larger samples do exist, they often group together cases that differ from one another in theoretically important ways.' Complexity in the form of multiple causation and equifinality can also make simple statistical comparisons misleading. But it is hard to elaborate more sophisticated statistical tests until one has a deeper baseline understanding of the nature of the phenomenon under investigation, as well as the categories and variables that make up candidate causes (Geddes, 1990: 131-50; Lustick, 1996: 505-18; Jervis, 1997). Wars — to continue with the same example — are similar to chemical and nuclear reactions in that they have underlying and immediate causes. Even when all the underlying conditions are present, these processes generally require a catalyst to begin. Chain reactions are triggered by the decay of atomic nuclei. Some of the neutrons they emit strike other nuclei prompting them to fission and emit more neutrons, which strike still more nuclei. Physicists can calculate how many kilograms of Uranium 235 or Plutonium at given pressures are necessary to produce a chain reaction. They can take it for granted that if a 'critical mass' is achieved, a chain reaction will follow. This is because trillions of atoms are present, and at any given moment enough of them will decay to provide the neutrons needed to start thereaction. In a large enough sample, catalysts will be present in a statistical sense. Wars involve relatively few actors. Unlike the weak force responsible for nuclear decay, their catalysts are probably not inherent properties of the units. Catalysts may or may not be present, and their potentially random distribution relative to underlying causes makes it difficult to predict when or if an appropriate catalyst will occur. If in the course of time underlying conditions change, reducing basic incentives for one or more parties to use force, catalysts that would have triggered war will no longer do so. This uncertain and evolving relationship between underlying and immediate causes makes point **prediction extraordinarily difficult**. It also makes more general statements about the causation of war problematic, since we have no way of knowing what wars would have occurred in the presence of appropriate catalysts. It is probably impossible to define the universe of would-be wars or to construct a representative sample of them. Statistical inference requires knowledge about the state of independence of cases, but in a practical sense that knowledge is often impossible to obtain in the analysis of international relations. Molecules do not learn from experience. People do, or think they do. Relationships among cases exist in the minds of decision-makers, which makes it very hard to access that information reliably and for more than just a very small number of cases. We know that expectations and behavior are influenced by experience, one's own and others. The deterrence strategies pursued by the United States throughout much of the Cold War were one kind of response to the failure of appeasement to prevent World War II. Appeasement was at least in part a reaction to the belief of British leaders that the deterrent policies pursued by the continental powers earlier in the century had helped to provoke World War I. Neither appeasement nor deterrence can be explained without understanding the context in which they were formulated; that context is ultimately a set of mental constructs. We have descriptive terms like 'chain reaction' or 'contagion effect' to describe these patterns, and hazard analysis among other techniques in statistics to measure their strength. But neither explains how and why these patterns emerge and persist. The broader point is that the relationship between human beings and their environment is not nearly so reactive as with inanimate objects. Social relations are not clock-like because the values and behavioral repertories of actors are not fixed; people have memories, learn from experience and undergo shifts in the vocabulary they use to construct reality. Law-like relationships — even if they existed — could not explain the most interesting social outcomes, since these are precisely the outcomes about which actors have the most incentive to learn and adapt their behavior. Any regularities would be `soft'; they would be the outcome of processes that areembedded in history and have a short half-life. They would decay quickly because of the memories, creative searching and learning by political leaders. Ironically, the `findings' of social science contribute to this decay (Weber, 1969; Almond and Genco, 1977: 496-522; Gunnell, 1982: Ch. 2; Ball, 1987: Ch. 4; Kratochwil, 1989; Rorty, 1989; Hollis, 1994: Ch. 9). Beyond these conceptual and empirical difficulties lies a familiar but fundamental difference of purpose. Boyle's Law, half-lives, or any other scientific principle based on probability, says nothing about the behavior of single units such as molecules. For many theoretical and practical purposes this is adequate. But social science ultimately aspires — or should aspire —to provide insight into practical world problems that are generally part of a small or very small n. In international relations, the dynamicsa nd outcomes of single cases are often much more important than any statistical regularities. The conception of causality on which deductive-nomological models are based, in classical physics as well as social science, **requires empirical invariance** under specified boundary conditions. The standard form of such a statement is this — given A, B and C, if X then (not) Y.4 This kind of bounded invariance can be found in closed systems. Open systems can be influenced by external stimuli, and their structure and causal mechanisms evolve as a result. Rules that describe the functioning of an open system at time T do not necessarily do so at T + 1 or T + 2. The boundary conditions may have changed, rendering the statement irrelevant. Another axiomaticcondition may have been added, and the outcome subject to multiple conjunctural causation. There is no way to know this a priori from the causal statement itself. Nor will complete knowledge (if it were possible) about the system at time T necessarily allow us to project its future course of development. In a practical sense, all social systems (and many physical and biological systems) are open. Empirical invariance does not exist in such systems, andseemingly probabilistic invariances may be causally unrelated (Harre and Secord, 1973; Bhaskar, 1979; Collier, 1994; Patomaki, 1996; Jervis, 1997). As physicists readily admit, prediction in open systems, especially non-linear ones, is difficult, and often impossible. The risk in saying that social scientists can 'predict' the value of variables in past history is that the value of these variables is already known to us, and thus we are not really making predictions. Rather, we are trying to convince each other of the logic that connects a statement of theory to an expectation about the value of a variable that derives from that theory. As long as we can establish the parameters within which the theoretical statement is valid, which is a prerequisite of generating expectations in any case, this 'theory-testing' or 'evaluating' activity is not different in a logical sense when done in past or future time.5

**Nobody can predict extinction**

**Yudkowsky ‘6** (Eliezer; Research Fellow at the Singularity Institute for Artificial Intelligence “Cognitive biases potentially affecting judgment of global risks” Forthcoming in Global Catastrophic Risks, eds. Nick Bostrom and Milan Cirkovic 8/31/06) MG

Any existential risk evokes problems that it shares with all other existential risks, *in addition to* the domain-specific expertise required for the *specific* existential risk. Someone on the physics-disaster committee should know what the term "existential risk" means; should possess whatever skills the field of existential risk management has accumulated or borrowed. For maximum safety, that person should also be a physicist. **The domain-specific expertise and the expertise pertaining to existential risks should combine in one person. I am skeptical that a scholar of heuristics and biases, unable to read physics equations,** could check the work of physicists who knew nothing of heuristics and biases. Once upon a time I made up overly detailed scenarios, without realizing that *every* additional detail was an extra burden. Once upon a time I really did think that I could say there was a ninety percent chance of Artificial Intelligence being developed between 2005 and 2025, with the peak in 2018. This statement now seems to me like complete gibberish. Why did I *ever* think I could generate a tight probability distribution over a problem like that? Where did I even get those numbers in the first place? Skilled practitioners of, say, molecular nanotechnology or Artificial Intelligence, will not automatically know the *additional* skills needed to address the existential risks of their profession. No one told me, when I addressed myself to the challenge of Artificial Intelligence, that it was needful for such a person as myself to study heuristics and biases. I don't remember why I first ran across an account of heuristics and biases, but I remember that it was a description of an overconfidence result - a casual description, online, with no references. I was so incredulous that I contacted the author to ask if this was a real experimental result. (He referred me to the edited volume *Judgment Under Uncertainty*.) I should not have had to stumble across that reference by accident. Someone should have warned me, as I am warning you, that this is knowledge needful to a student of existential risk. There should be a curriculum for people like ourselves; **a list of skills we need *in addition to* our domain-specific knowledge.** I am not a physicist, but I know a little - probably not enough - about the history of errors in physics, and a biologist thinking about superviruses should know it too. I once met a lawyer who had made up his own theory of physics. I said to the lawyer: You cannot invent your own physics theories without knowing math and studying for years; physics is hard. He replied: But if you really understand physics you can explain it to your grandmother, Richard Feynman told me so. And I said to him: "Would you advise a friend to argue his own court case?" At this he fell silent. He knew abstractly that physics was difficult, but I think it had honestly never occurred to him that physics might be as difficult as lawyering.

**Util Bad**

**We are only responsible for the action of the plan.**

**Harris ‘8** (Alex; 8/15/8; J.D. Stanford University, B.A. from Harvard University, Practicing Appellate and Constitutional Law at Gibson, Dunn & Crutcher LLP, former Adjunct Analyst at The Competitive Enterprise Institute; Philosopher’s Corner, “Philosopher's Corner: The Principle of Intervening Action,” <https://cei.org/blog/philosophers-corner-principle-intervening-action)> \*PIA = Principle of Intervening Action\*

Gewirth takes the position that we are **solely responsible** for the morality of **our own actions** in two senses. First, only we are responsible for the acts we commit, **even if someone else's action** caused us to act as we did. (For example, if a woman's husband cheated on her and she, upon finding out, grew enraged and killed his lover, she - not he - would bear sole responsibility.) Second, we are only responsible for our own actions, **even if they lead** to other actions. Thus, we have a **preeminent duty** to never act immorally, even if doing so would preclude others from taking even more immoral actions. Gewirth contends that never violating the negative rights of another "is an obligation so fundamental that it **cannot be overridden** even to prevent evil consequences from befalling some persons." He clarifies with an example. Imagine that a group of terrorists kidnaps a woman and offers her son a choice: he must torture his mother or they will blow up a city with a nuclear weapon. Gewirth argues that the son has a primary duty to not violate the rights of his mother, whereas he is not the actor who is blowing up the city - the terrorists are the moral agents responsible for that action, not the son. If the son had the choice, he would pick neither. His duty is to never violate rights; the only way to fulfill this is to not torture his mother. Gewirth argues: "It would be unjustified to violate the mother's right to life in order to protect the rights to life of the many other residents of the city. For rights **cannot be justifiably protected** by violating another right." PIA is the only **consistent, justifiable** moral theory of consequences. First, one should note that only PIA sets a **non-arbitrary limit** on the string of effects that can factor into the moral calculation. PIA says that **no consequences** of other actions can count; the only other non-arbitrary standard says that all consequences in the chain must count. One cannot claim that I am responsible for only, say, the first four other actions resulting from my action. One must either consider only my actions or all resulting actions. Thus, if the destruction of the city by terrorists actually ended up preventing more rights violations by, say, staving off a Malthusian population crunch that would result in mass starvation and world war, then the consequentialist position has to **endorse the terrorists' action**. Consequentialists have to count **every effect** in the chain, even in the absurdly far-off future, to determine whether an action is moral. This fact, of course, does not by itself constitute a reason to reject consequentialism in favor of PIA, but it does suggest that PIA is the **only reasonable** interpretation of the requirement of non-consequentialism. It also suggests an implausible feature of consequentialism. I went on to demonstrate how the libertarian principle of self-ownership supports PIA and why people cannot be responsible for all effects of their actions: Since we are born owning ourselves and nothing else, controlling our mind and body and no one else's, it makes **perfect sense** that we should be responsible for only the actions that **we ourselves commit**. Some could argue that we should be responsible for the results of these actions. PIA states that **we are**. If a person gets a wrecking ball and knocks over a building, which then falls and crushes twenty people, the person is to some degree responsible for those **results**. But this is not the case if **someone else's action intervenes**, because another moral agent is the more **proximate cause** of the effects; she has stepped into the **line of causation** to take the moral responsibility. When you act upon a rock that you hurl at an enemy's face, you are responsible for the effects of the rock for two reasons: first, you are using force upon the rock; secondly, the rock has no agency over the effects it causes. The rock, by the fact that it has no agency of its own, is merely your tool, an extension of your agency. But neither of these reasons holds for using non-coercive measures that result in a person's action. As long as one does not use coercion to compel another to commit a rights-violating action, one has not reduced that other person's agency. Possessing full agency, the person is **morally responsible** for the **totality** of her actions; thus no one else can assume **any portion** of that responsibility. You are not responsible for anyone else's free actions and no one else is responsible for yours. If the son were somehow partially responsible for the terrorists blowing up the city, that would necessarily diminish, by whatever fraction of responsibility the son assumed, the terrorists' responsibility for that action. They would not be wholly responsible, because the son had caused their action. But this must not be the case; the terrorists must be held totally responsible for the destruction of the city. Consequentialists ask, "Which set of rights-violations do you endorse: the torture of the mother, or the deaths of the millions?" Gewirth responds that PIA endorses **neither**. PIA gives the terrorists **complete responsibility** for their actions, and emphatically condemns them, in a way that **no other position** is capable of. Only PIA is capable of giving rights their **supreme status** by proclaiming that they may never be violated for any reason, including preventing future rights-violations.

**Utilitarian balancing can’t justify continued violence. One-percent risk of a solvency deficit to a counterplan infinitely outweighs even the largest net-benefit.**

**Gross 1** (James A. Gross, Professor of Industrial and Labor Relations at Cornell University, holds a Ph.D. from the University of Wisconsin-Madison, 2001, “A Human Rights Perspective on U.S. Education: Only Some Children Matter,” *The Catholic University Law Review* (50 Cath. U.L. Rev. 919), Summer, Available Online to Subscribing Institutions via Lexis-Nexis)

VI. Concluding Observations

To understand that education is a human right is to understand that the problems of education in this country and the proposed solutions are **inextricably interconnected** with issues of morality, justice and values. Fundamental issues of human rights, justice and morality must be addressed and resolved before any reconstruction of the educational system is attempted. What is excused as misfortune must be **recognized as injustice** and what has been dismissed as the status quo must be traced to the action or inaction of the unjust.

A just society, particularly one with the economic resources of the United States, would not choose to reject any of its children. A just society would treat each of its children as an "unprecedented wonder" n243 and would be committed to enabling them to realize their potential for living a full human life. n244 Each child would be recognized for the person he is; his presence on this earth would be treated as an "unconditional blessing." n245

This recognition and celebration of life is the core principle of human rights. It was recognized by a Freedmen's Bureau commissioner who urged that the freed people in the Bureau's schools be "treated as men with immortal souls rather than as beasts of burden or machines for pulling cotton." n246 More than 100 years later, Thomas Sowell similarly noted that the "only common denominator among the successful schools [in the black community/ghetto] was that the students were treated like human beings and everything was geared to the expectation that they would succeed." n247 The children understood that they were important in and of themselves. n248

[\*952] Conscious choices violate the human rights of certain children. Yet human rights constitute **the most essential moral claims** that all human beings can assert. n249 They confirm the sacredness of human beings and their **intrinsic dignity**. Human rights are entitlements. The great disparity in the amount of money spent for some compared to that spent on the education of other people's children is a measure of **how little certain children are valued as human beings**. As a result, a message is sent that those children "deserve to be neglected [and] to be surrounded by a blatant lack of respect." n250

A solution to this problem will require the problem solvers to know what it is like for children to grow up rejected and shunned by the dominant society, what it means and does to them, and whether they think they deserve to be treated that way. As Kozol asks, "what is it that enables some of them to pray? When they pray, what do they say to God?" n251 Other previously ignored questions must also be answered:

how certain people hold up under terrible ordeal, how many more do not, how human beings devalue other people's lives, how numbness and destructiveness are universalized, how human pity is at length extinguished and the shunning of the vulnerable can come in time to be perceived as natural behavior ... . How does a nation deal with those whom it has cursed? n252

Others wonder about the impact of long-standing devaluation on both the children devalued and on those responsible for that devaluation: "after all that has happened, in history and in our own time, can black people still be seen with empathy and without sentimentality as human beings with aspirations and potential that deserve fulfillment?" n253 Andrew Hacker maintains that persuading Americans to care about children other than their own is imperative because indigent children are looked upon as a burden. n254

Where is the public indignation at the abuse of innocent children who have done nothing wrong? Despite a "reverence for fair play" and a "genuine distaste for loaded dice" in the United States, Kozol maintains [\*953] that in the realms of education, health care and inheritance of wealth, fairness is not evident. n255 In those areas, Kozol says, "we want the game to be unfair and we have made it so; and it will likely so remain." n256 If our motives can be judged most accurately by our actions or inaction, Hacker and Kozol's perceptions are on the mark. Many in our country, including children, are isolated in helplessness while others choose to isolate themselves by their own selfishness. It is a selfishness that consists not only of an unwillingness to redistribute resources to others in need, but also of a **deliberate perpetuation of an unfair distribution of the benefits of the educational system** which secures advantages in society.

Americans pride themselves on their morality. The "American Creed" is the ideological foundation of the nation, encompassing the ideals of the inherent dignity of the individual human being, and of the fundamental equality of all, as well as "inalienable" rights to freedom, justice and fair opportunity. All of these ideals are reconciled within the framework of the common good. These are the elements of a democratic creed that, although pre-dating the United States, represents the "national conscience." n257 The creed is the basis for the realization of the "American Dream," which in addition to being a dream of wealth has also "been a dream of being able to grow to fullest development as a man and woman" to benefit "the simple human being of any and every class." n258

In 1944, Swedish economist Gunnar Myrdal characterized U.S. race relations as an "American Dilemma": the moral dilemma of the disparity between ideals and actual behavior. n259 It is tragic that any such gap remains after all of these years. Yet, it is not unrealistic to believe in and work for change unless those with economic and political influence are completely hypocritical. The civil rights and women's rights movements in this country are among the precedents that justify some optimism and hope.

No matter how discouraging the prospects for fundamental change in the educational system, it would be even more **irresponsible to fail to act**. n260 If human rights violations are to end, then the moral choices that [\*954] underlie those violations and the values that influence those moral choices must be changed. n261 Without that change, we will continue merely to remodel on a faulty foundation. Despite commentaries about the futility of trying to reverse these choices, fundamental change **is** possible and one of the many reasons for that change is the ability of challengers to redefine a policy issue.

Acceptance of education as a human right changes our understanding of the essential purpose of education and requires a fundamental and thorough redefinition of education policy. The primary objective of education policy would become compliance with the rights of all children to the type and quality of education needed to live full human lives **rather than**, as now, **conceiving of education as merely a utilitarian instrument for maximizing payoff for those who invest in it** - or for those who can afford the type of education most likely to provide the greatest return on investment. It puts into sharp historical and cultural perspective the fact that since the Universal Declaration of Human Rights in 1945, nations from all over the world have recognized education as a human right while our own Supreme Court does not consider education to be even a constitutional right.

It may be that domestic human rights issues go unacknowledged by the public because of the myth that the United States is a paragon of human rights observance. As human rights become more important in international relations, this country is vulnerable to charges of hypocrisy for attempting to maintain a "facade of championing human rights when it does not protect the rights of its own citizens." n262 Despite the rhetoric about the sanctity of human rights, hypocritical or not, it is likely that most people in this country comprehend human rights only in the context of such egregious human evils as genocide or systematic torture. Beyond that there is little understanding of the meaning, significance and implications of human rights.

All education systems want to produce a certain kind of human being, and values have always been an essential and unavoidable part of education. Ironically, therefore, the redefinition of education policy [\*955] issue requires education. From the time they start school, children need to learn about human rights and to respect the human rights and dignity of all people regardless of race, color, language, gender, or faith. Human rights education needs to occur at all levels from elementary school through college or university.

Promotion of internationally recognized human rights principles emerging in international law, moreover, would educate our judiciary as well as the public. These international human rights principles pose a growing challenge to what some experts consider the isolation and provincialism of U.S. courts. n263 Given the influence of values on judicial decision-making, these human rights principles provide an important source of law for U.S. courts to use in the interpretation of the Constitution, including filling in the gaps in constitutional protections. To ignore those principles is to express indifference to them and expresses a willingness to put the United States in direct conflict with international law. n264

No attempt is made here to spell out the details of a curriculum or the content of specific course subjects needed to enable people to live full human lives. However, a quality education is about reading, writing, computing, communicating, imagining, thinking, reasoning, creating, participating, questioning, analyzing, challenging, judging, and changing. It is about the unprecedented wonder of each and every human being, the rights and duties of each other. It is about history and heritage as well as partaking in cultural stories and heritage. It is about sharing all the intellectual adventures at the heart of civilization. It is about morals and ethics and the content of character. It is also about participating in decisions that affect one's life.

A quality education must not be indoctrination in an "Aren't-We-Americans-Just-Dandy curriculum" as Theodore and Nancy Sizer called it. n265 Education needs to have a global perspective with an understanding of all peoples, their cultural heritage, values, problems and ways of life. Education needs to be about human solidarity, respect for human dignity, the equal rights of all human beings, and justice and equality for all people.

There is **no reason** that can justify the perpetuation of human rights violations to education: not transparent appeals to the democratic principle of local control of education (it would be a perverted [\*956] democracy that commits or tolerates violations of the human rights of children); not a state's use of local control as an excuse rather than as a justification for interdistrict inequality; n266 and not the federal government's evasion of the duty by hiding behind the myth that education is exclusively a state and local matter in this country. n267

**A just society would not tolerate anything less than the end of these violations of our children's human right to education**. Of course, our willingness to end these violations depends on **the type of a society we desire** and **what kind of people we want to be**.

**This logic of utilitarianism culminates in extinction**

**Santos 3** 2003, Boaventura de Souza Santos is a Professor of Sociology at the University of Coimbra, “Collective Suicide?”, Bad Subjects, Issue # 63 , http://www.ces.fe.uc.pt/opiniao/bss/072en.php

According to Franz Hinkelammert, the West has repeatedly been under the illusion that it should **try to save humanity by destroying part of it**. This is a salvific and sacrificial destruction, committed in the **name of the need to radically materialize all the possibilities opened up by a given social and political reality** over which it is supposed to have total power. This is how it was in **colonialism**, with the **genocide of indigenous peoples**, and **the African slaves**. This is how it was in the **period of imperialist struggle**s, which caused **millions of deaths** in two world wars and many other colonial wars. This is how it was in **Stalinism**, with the **Gulag** and in **Nazism**, with the **holocaust**. And now today, this is how it is in neoliberalism, with the **collective sacrifice of the periphery** and **even the semiperiphery of the world system.** With the war against Iraq, it is fitting to ask whether what is in progress is a new genocidal and sacrificial illusion, and what its scope might be. It is above all appropriate to ask if the new illusion will not herald the radicalization and the ultimate perversion of the western illusion: **destroying all of humanity in the illusion of saving it.** Sacrificial genocide arises from a totalitarian illusion that is manifested in the belief that there are no alternatives to the present-day reality and that the problems and **difficulties confronting it arise from failing to take its logic of development to its ultimate consequences.** If there is unemployment, hunger and death in the Third World, this is not the result of market failures; instead, it is the outcome of the market laws not having been fully applied. If there is terrorism, this is not due to the violence of the conditions that generate it; it is due, rather, to the fact that total violence has not been employed to physically eradicate all terrorists and potential terrorists. **This political logic is based on the supposition of total power and knowledge**, and on the radical rejection of alternatives; it is ultra-conservative in that it aims to infinitely reproduce the status quo. Inherent to it is the notion of the end of history. During the last hundred years, the West has experienced three versions of this logic, and, therefore, seen three versions of the end of history: Stalinism, with its logic of insuperable efficiency of the plan; Nazism, with its logic of racial superiority; and neoliberalism, with its logic of insuperable efficiency of the market. The first two periods involved the destruction of democracy. The last one trivializes democracy, disarming it in the face of social actors sufficiently powerful to be able to privatize the State and international institutions in their favour. I have described this situation as a combination of political democracy and social fascism. One current manifestation of this combination resides in the fact that intensely strong public opinion, worldwide, against the war is found to be incapable of halting the war machine set in motion by supposedly democratic rulers. At all these moments, a **death drive**, a catastrophic heroism, predominates, **the idea of a looming collective suicide**, **only preventable by the massive destruction of the other**. Paradoxically, the broader the definition of the other and the efficacy of its destruction, the **more likely collective suicide becomes**. In its sacrificial genocide version, neoliberalism is a mixture of market radicalization, neoconservatism and Christian fundamentalism. **Its death drive takes a number of forms, from the idea of "discardable populations"**, referring to citizens of the Third World not capable of being exploited as workers and consumers, to **the concept of "collateral damage" ,** to refer to the deaths, as a result of war, of thousands of innocent civilians. The last, catastrophic heroism, is quite clear on two facts: according to reliable calculations by the Non-Governmental Organization MEDACT, in London, between 48 and 260 thousand civilians will die during the war and in the three months after (this is without there being civil war or a nuclear attack); the war will cost 100 billion dollars, enough to pay the health costs of the world's poorest countries for four years. Is it possible to fight this death drive? We must bear in mind that, historically, **sacrificial destruction has always been linked to the economic pillage of natural resources and the labor force, to the imperial design of radically changing the terms of economic, social, political and cultural exchanges in the face of falling efficiency rates postulated by the maximalist logic of the totalitarian illusion in operation.** It is as though hegemonic powers, both when they are on the rise and when they are in decline, repeatedly go through times of primitive accumulation, legitimizing the most shameful violence in the name of futures where, by definition, **there is no room for what must be destroyed**. In today's version, the **period of primitive accumulation consists of combining neoliberal economic globalization with the globalization of war. The machine of democracy and liberty turns into a machine of horror and destruction.**

1. [↑](#footnote-ref-1)
2. [↑](#footnote-ref-2)
3. [↑](#footnote-ref-3)