# 1AC

### Advantage

#### Insulin is prohibitively expensive – new insulin analogues move the needle from human insulin to a lower quality, more expensive drug

Peccoud et al 18 Jenna E. Gallegos [],1 Christopher Boyer,2 Eleanore Pauwels,3 Warren A. Kaplan,4 and Jean Peccoud [Prof. Jean Peccoud joined the department in January 2016 as the Abell chair in synthetic biology]1,\*, December 18, “The Open Insulin Project: A Case Study for ‘Biohacked’ Medicines””, Trends in Biotechnology Vol 36 No. 12, <https://www.cell.com/trends/biotechnology/pdf/S0167-7799(18)30200-2.pdf> DD AG

Since its discovery in 1921, insulin has revolutionized the quality and quantity of life for persons with diabetes. Yet, despite its long market history, the cost of insulin has continued to rise. For example, insulin prices tripled between 2002 and 2013 [9], costing uninsured patients as much as US$400 per month [10]. In inner cities, the leading cause of diabetic ketoacidosis – a potentially fatal condition – is stopping or inconsistent insulin treatment, and cost is a major reason reported for this [11]. Cited examples of health risks from high insulin costs include rationing treatments, using expired products, fasting, and even intentionally inducing diabetic ketoacidosis in order to obtain insulin from hospital emergency rooms [12,13]. While many other lifesaving medications have become available as less expensive generics, the high price of insulin is maintained in part by the small number of multinational corporations that dominate the insulin market and the complex and opaque pricing and supply chain [14].

The structure of human insulin is not patent protected, but the market has shifted to the production of genetically modified insulin analogues, in large part because the pharmaceutical industry has seen fit to incrementally innovate, raise the price, and phase out the old forms of insulin [10,14]. Insulin analogues are marketed as having additional benefits such as fast or long-acting properties and labeling for pediatric or pregnant patients. However, many experts argue that the originally approved human insulin is just as effective for most patients [15,16], so it is difficult to say whether patients who, because of lack of insurance and/or socio-economic inequalities [17,18], should be literally paying the price for insulin analogues when human insulin may well be as effective.

Only now, with intellectual property (i.e., patents) for many insulin analogues having recently expired or expiring soon [19], have biosimilar insulin analogues been marketed. However, there is still no inexpensive supply of insulin biosimilars for people living with diabetes in North America, and Americans are paying a steep price for the ‘continued rejuvenation’ of this medicine [10]. Meanwhile, at least 11 insulin biosimilars are marketed (under less stringent regulatory frameworks) at considerably lower price points in China, India, Mexico, Pakistan, Peru, and Thailand [20]. Studies comparing a handful of these biosimilars to innovator insulins showed no meaningful differences [19,20].

It is difficult for potential biosimilar manufacturers to compete in the US because the regulatory system explicitly favors existing manufacturers. First, the main purpose of clinical trials is to establish similarity to an innovator biologic, not clinical benefit per se [21,22]. This emphasis on proof-of-similarity strongly favors the pharmaceutical companies that produced the original as only they have access to the confidential manufacturing protocols.

Additionally, while competitors wishing to manufacture a biosimilar are subject to strict regulatory oversight, changes by existing manufacturers rarely require clinical trials, and the resulting biosimilar is treated as interchangeable [23]. This discrepancy is deemed excusable because a manufacturer that modifies its own processes is supposed to have extensive knowledge and information about the product. It is thus no surprise that the first insulin biosimilar approved in the US, Basaglar1 (Box 2), was produced by Eli Lilly, which already owned 20% of the market share for insulin [24].

While generic drugs are typically 80% less expensive than the equivalent name-brand medications, Basaglar1 is only 15% cheaper than the innovator biologic Lantus1 [25]. The minimal cost saving associated with biosimilar insulins likely has little to do with manufacturing cost; the market value of pharmaceutical insulin is over $1000 per gram [9], while insulin costs roughly $50–75 per gram to manufacture [24]. Instead, costs are largely set by the intellectual property holders in response to the complex regulatory environment surrounding biologic drugs. Developers of biohacked insulin will thus have to navigate both intellectual property and regulatory hurdles in order to develop a more affordable model for insulin production.

#### IP perpetuates evergreening, specifically for insulin – that prevents the creation of cheap, generic medicine

Greene 15 Jeremy A. Greene, M.D., Ph.D [I received an MA in medical anthropology from Harvard in 2004, the MD and PhD degrees in the history of science from Harvard in 2005]., and Kevin R. Riggs, M.D., M.P.H., March 19, 2015, “Why Is There No Generic Insulin? Historical Origins of a Modern Problem”, New England Journal of Medicine 372:1171-1175, <https://www.nejm.org/doi/full/10.1056/NEJMms1411398> DD AG

Reducing the problem of generic insulin to the contemporary debate over biosimilarity ignores the historical reason why we have always lacked generic insulin: incremental innovation has repeatedly precluded the formation of a generic-insulin industry in North America when earlier patents expired. The history of insulin hasn't followed the standard chronology of pharmaceutical innovation, in which patent monopolies predictably give way to generic competition.

Viewed in historical perspective, insulin is not a single entity but a family of related products that has evolved through incremental improvements. Subsequent iterations of insulin represented actual innovations, each one being safer, more effective, or more convenient than its predecessor. And yet after generations of incremental innovation, insulin may be no more affordable than it was when the original patent holders sold their stake for $1 to ensure access to this essential medicine.

Pharmaceutical-industry analysts have described a repatenting tactic called evergreening, in which a series of related patents — often on metabolites or optical isomers — extend the life of a product after initial patent expiration.23 Evergreening can shift market share within a family of products: for example, after Pfizer lost patent exclusivity on the antiepileptic agent gabapentin (Neurontin) in 2004, it retained a healthy share of the market through patents on a metabolic cognate, pregabalin (Lyrica). Critics of evergreening often claim that the incremental innovations leading from a given drug to a “me-too” drug are trivial: pregabalin, for example, is not clearly safer or more efficacious than gabapentin.

But the cascading generations of insulin products can hardly be dismissed as simply “me-too” medicines. Protamine insulin offered a distinct advantage over regular insulin, NPH insulin offered a distinct advantage over protamine insulin, and so on. On the whole, insulin today is demonstrably safer and more convenient to use than products available in 1923. But whether each incremental innovation is worth the price we pay, in a world where insulin remains unaffordable to many patients with diabetes, is less certain. When lente insulin was introduced in the 1950s, some observers questioned whether its minimal theoretical advantages over NPH warranted the complexity introduced by adding another insulin formulation to the market.24 The theoretical advantages offered by the monocomponent extract insulins may sometimes have been outweighed by the inconvenience and risk caused by transitioning patients to an insulin of different potency.25 Although recombinant insulin was heavily advertised as a clinically superior agent in the 1980s (Figure 1), almost no evidence was provided to demonstrate its superiority to the best available animal-extract insulins.26 Although long-acting analogues cause less hypoglycemia than NPH does,27 it has yet to be shown that analogues lead to better long-term outcomes than standard recombinant human insulin does.28

No doubt for many patients, these incremental innovations were worth the added price. What's surprising is that the trailing edge of old insulin products did not generate a market for generic competition but rather became a set of obsolete products that were promptly removed from the U.S. market. Pork and beef insulins are not merely underutilized, they are unavailable for human use in the United States. Even when practitioners prescribe NPH and R insulin in place of insulin glargine and insulin aspart, these cheaper prescriptions are filled with newer recombinant products sold under brand names. And yet on the whole, it's hard to say that contemporary patients who cannot afford their insulin (let alone the patent-protected glucometers and test strips required to adjust the dose) are well served by having as their only option an agent that is marginally more effective than those that could have been generically available 50 or 30 or 10 years ago, had generics manufacturers introduced cheaper versions when patents expired.

Generic-drug companies have evidently not considered it worthwhile to invest in the additional good manufacturing practices needed to produce a version of insulin that may already be obsolete, when off-patent small-molecule drugs represented lower-hanging fruit. Only recently, with insulin-analogue patents expiring and no next-generation products on the horizon, have prominent generics manufacturers shown serious interest in the insulin market.

It is hard to overstate the economic and public health impact of generic drugs in improving access to safe, effective, inexpensive medications in the United States. In the early 1960s, fewer than 1 in 10 medicines dispensed in pharmacies were generic, and most prescription drugs were effectively monopolies. Today, more than 80% of prescriptions are filled with generics, which saves the health care system billions of dollars each year.29,30 These savings are critical both for payers that are squeezed by rising health care costs and for patients, because lower medication costs are associated with better adherence31 and better outcomes.32

But the case of insulin demonstrates that the generics market is like other markets — not an automatic phase in the life cycle of a drug. As the increasing waves of generic-drug shortages in the past decade also remind us, the drugs that ultimately see extensive generic competition differ from those that attract few, if any, manufacturers. The history of insulin highlights the limits of generic competition as a public health framework. Nearly a century after its discovery, there is still no inexpensive supply of insulin for people living with diabetes in North America, and Americans are paying a steep price for the continued rejuvenation of this oldest of modern medicines.

#### This isn’t just a one off – it’s been happening for the last century

Peccoud 18 Jean Peccoud [Prof. Jean Peccoud joined the department in January 2016 as the Abell chair in synthetic biology.], 9-13-2018, "After a century, insulin is still expensive – could DIYers change that?," Conversation, <https://theconversation.com/after-a-century-insulin-is-still-expensive-could-diyers-change-that-99822> DD AG

Soon after Federick Banting discovered that insulin could be used to treat diabetes in 1921, he sold the patent to the University of Toronto for about a dollar. Banting received the Nobel prize because his discovery meant a life-saving drug could become widely available. Nearly a century later, an American with diabetes can pay as much as US$400 per month for insulin, driving some uninsured patients to desperate and dangerous measures. Clearly, something went wrong.

Our lab studies biosecurity, so when we heard that a group of do-it-yourself biologists was working to solve the insulin affordability problem by figuring out how to manufacture insulin patent-free, we got to know them. After digging into the insulin affordability issue, we argue that what’s keeping insulin expensive is not patents – it’s regulations. By operating in a regulatory blind spot, DIYers could upset the status quo for drug production.

Discovering and developing drugs is expensive. Patents help drug companies recoup the costs from their investments by granting them a monopoly for a limited time. Once the patent expires, competing companies can begin producing generics: off-brand versions of a patented drug. This healthy competition drives prices down.

So why, with the original patent long-expired, is there still no affordable generic insulin?

The insulin for purchase today is not the same insulin used to treat diabetic patients nearly 100 years ago. That insulin came primarily from animals. Today, insulin is brewed up by microbes that have been genetically engineered with the gene for human insulin.

And insulin is seldom injected with an old-fashioned syringe and needle anymore. Now there are insulin pens, pumps, test strips and other devices that improve the quality of life for diabetic patients. Pharmaceutical companies have also modified the chemical formula to produce faster-acting or longer-lasting insulins.

With each of these inventions came a new patent.

But the benefits of these “improved” insulins are debatable, and there’s nothing preventing competing companies from selling older, long off-patent versions of insulin. So what’s the holdup?

Regulations keep insulin expensive

Insulin is a biologic drug, which means it’s produced by a living organism, not a chemical reaction. This process, called biomanufacturing, is more inconsistent than chemical synthesis of non-biologic drugs like aspirin.

Making reliable biologic drugs is a little like winemaking. Even though the winemaker carefully follows a well-established process, minute differences will affect the final product. It’s always wine, but some vintages are better than others and tasting the wine is the only way to evaluate the final product.

So if a new company wants to make insulin, that insulin has to be tested on patients in expensive clinical trials. Bringing a biologic drug to market can cost as much as $250 million. No company can afford that lump if it can’t file for a patent to recoup the investments.

That’s why there’s only one “generic” insulin available so far. It’s made by a company that was already a major player in the insulin market, and it’s only 15 percent cheaper than the patented version. By comparison, most non-biologic generic drugs cost 80 percent less than the original.

Obviously, regulations are important for keeping insulin safe, but at what cost? Ten percent of people living with diabetes in the U.S. are uninsured, and there are nearly 10,000 crowdfunding campaigns related to insulin on the site GoFundMe alone. Stories about diabetic patients ending up hospitalized or worse because they tried to ration their insulin are all-too common.

#### Thus the plan, the United States of America ought to reduce intellectual property protections for insulin.

#### The plan would be implemented through the Affordable Drug Manufacturing Act – precedent makes it normal means

Scott 18 Dylan Scott [grew up in Ohio, lived in Las Vegas for a year and moved to Washington in 2011. I cover health care and other domestic policy.], 12-20-2018, "Elizabeth Warren’s ambitious new bill to lower generic drug prices, explained," Vox, <https://www.vox.com/policy-and-politics/2018/12/20/18146993/elizabeth-warren-2020-election-drug-prices-bill> DD AG  
The two lawmakers are targeting generic drugs — knockoff versions of brand-name medications that have lost their patent protections — specifically in their legislation. Generic drugs are supposed to lower drug prices by introducing cheaper alternatives to the brand-name version. But the generic drug industry has come under a lot of scrutiny in recent years, both for hiking generic prices and becoming the target of a historic lawsuit for anti-competitive practices. The system is often not working as intended.

Here’s what Warren and Schakowsky would propose to do: Under limited circumstances, the federal government would produce a more affordable generic version of certain drugs. These are the scenarios when the feds could start manufacturing their own medications:

* If no company is producing a generic version of the drug
* If only one or two companies are producing a drug and there is either a price hike or a drug shortage
* If only one or two companies are producing a drug, the price makes it difficult for some patients to afford, and the World Health Organization classifies it as an “essential medicine”

The bill allows the federal government to either produce the drugs itself or contract an outside company to do it. It would set “fair” prices to cover the costs of making the drugs. They also want the federal government to start producing insulin, which helps treat diabetes (which tens of millions of Americans have) and has seen its prices triple or so over the past decade.

“This proposal could be a helpful intervention for those generic drugs that have recently seen price spikes or are in shortage,” Rachel Sachs, who follows the drug pricing debate at Washington University in St. Louis, told me.

I heard the same from others — “an intriguing idea” as Craig Garthwaite, a health economist at Northwestern University, said.

Major hospitals had the same idea. They’ve set up a new nonprofit drug company, along with some philanthropic groups, that would produce drugs under similar circumstances as the Warren-Schakowsky bill. This is one of the hot ideas in health care right now.

There are some questions, of course, starting first and foremost with how well equipped the government is to make medications. The Food and Drug Administration previously forced the National Institutes of Health to shut down its drug manufacturing facilities over quality concerns — though this legislation notably allows the federal government to contract with private companies to do the actual drug producing.

#### IP stands in the way of innovative, biohacked insulin that solves diabetes and is cheaper than existing medicines – only the plan allow for a new wave of biohacked innovation – turns the innovation DA

Peccoud et al 18 Jenna E. Gallegos [],1 Christopher Boyer,2 Eleanore Pauwels,3 Warren A. Kaplan,4 and Jean Peccoud [Prof. Jean Peccoud joined the department in January 2016 as the Abell chair in synthetic biology]1,\*, December 18, “The Open Insulin Project: A Case Study for ‘Biohacked’ Medicines””, Trends in Biotechnology Vol 36 No. 12, <https://www.cell.com/trends/biotechnology/pdf/S0167-7799(18)30200-2.pdf> DD AG

Biohacked insulin will face different intellectual property barriers depending on the distribution model. If the Open Insulin Project succeeds in developing and releasing a protocol for insulin manufacturing, and that protocol is adapted for personal use (as epinephrine auto-injectors have been), intellectual property will likely not be a substantial obstacle. Personal use of ‘home-brewed insulin’ would not trigger any patent considerations in most European countries, as the exclusive exploitation rights granted by a patent are restricted to commercial exploitation. In Europe, a private person who builds a patented invention and/or uses a patented method in her own home for her own personal goals generally cannot infringe on a patent. The reasoning behind this is that such a situation cannot harm the patent holder. In the US, the law is stricter, and it forbids anyone from making, using, or experimenting with an invention, even when the use is not commercial, except in very limited cases [26]. Practically speaking, however, since patent infringement lawsuits are very expensive, and it is Difficult to track restricted use in private, an individual would rarely, if ever, be prosecuted for using an invention in her own home. In the case of insulin, the safety ramifications of this scenario are obvious and will be discussed more thoroughly in the following sections.

Any other innovation ecosystem for insulin (e.g., ‘magistral’ production or technology transfer to a generic company) may run afoul of patents on the molecule and the production process, provided such patents exist. We note that there are plenty of patent applications filed for various ‘next generation’ insulin analogs, methods of making them, and methods of using them [27]. However, patents protecting the amino acid sequence of unmodified human insulin itself and of some recently off-patent insulin analogues are not a major barrier to the market introduction of affordable insulin. Patents protecting production methods of insulin are a more likely intellectual property obstacle.

Manufacturing is typically protected by a combination of patents and proprietary, non-patented know-how, or ‘trade secrets,’ which do not expire like patents. Manufacturing intellectual property includes the strain of microorganism used to biologically manufacture, or ‘express,’ the insulin and the specifics of the microbial fermentation process and recovery/purification of the expressed protein. Trade secrets are often used to protect non-patentable information. An insulin ‘bio hacker,’ however, can independently uncover or stumble upon and ‘acquire’ a trade secret.

### Framework

#### The standard is decreasing structural inequalities

#### Prefer:

#### 1] Prioritize structural impacts – worst-case scenario predictions are based on threat exaggeration – distorts rational decision-making and justify preemptive warfare

Mueller & Stewart ’11 [John, Woody Hayes National Security Studies and Professor of Political Science @ Ohio State University, Mark, Professor of Civil Engineering and Director of the Centre for Infrastructure Performance and Reliability at the University of Newcastle in Australia, “Terror, Security, and Money”, page numbers below]

Focusing on Worst-Case Scenarios Cass Sunstein, who seems to have invented the phrase "probability neglect," assesses the version of the phenomenon that comes into being when "emotions are intensely engaged." Under that circumstance, he argues, "people’s attention is focused on the bad outcome itself and they are inattentive to the fact that it is unlikely to occur." Moreover, they are inclined to "demand a substantial governmental response-even if the magnitude of the risk does not warrant the response." It may be this phenomenon that Treverton experienced. Playing to this demand, government officials are inclined to focus on worst-case scenarios, presumably in the knowledge, following Sunstein's insight, that this can emotionally justify just about any expenditure, no matter how unlikely the prospect the dire event will actually take place. Accordingly; there is a preoccupation with "low probability/ high consequence" events, such as the detonation of a sizable nuclear device in midtown Manhattan. The process could be seen in action in an article published in 2008 by Secretary of Homeland Security (DHS) Michael Chertoff. He felt called upon to respond to the observation that the number of people who die each year from international terrorism, while tragic, is actually exceedingly small. "This fails to consider," he pointed out, "the much greater loss of life that Weapons of mass destruction could wreak on the American people." That is, he was justifying his entire budget-only a limited portion of which is concerned with Weapons of mass destruction by the WMD threat, even while avoiding assessing its likelihood. It is sometimes argued that conventional risk analysis breaks down under extreme conditions because the risk is now a very large number (losses) multiplied by a very small number (attack probability). But it is not the risk analysis methodology that is at fault here, but our ability to use the information obtained from the analysis for decision making. A "high consequence" event has been defined to be a "disaster" or "catastrophe" resulting in "great human costs in life, property environmental damage, and future economic activity" However, depending on how one weighs the words in that definition, there may have been only one terrorist event in all of history that qualifies for inclusion. Moreover, the vast bulk of homeland security expenditures is not focused on events that fit a definition like that, but rather on comparatively low-consequence ones, like explosions set off by individual amateur jihadists. Analyst Bruce Schneier has written penetratingly of worst-case thinking. He points out that it , involves imagining the worst possible outcome and then acting as if it were a certainty. It substitutes imagination for thinking, speculation for risk analysis, and fear for reason. It fosters powerlessness and vulnerability and magnifies social [immobilization] ~~paralysis~~. And it makes us more vulnerable to the effects of terrorism. It leads to bad decision making because it's only half of the cost-benefit equation. Every decision has costs and benefits, risks and rewards. By speculating about what can possibly go wrong, and then acting as if that is likely to happen, worst-case thinking focuses only on the extreme but improbable risks and does a poor job at assessing outcomes. It also assumes "that a proponent of an action must prove that the nightmare scenario is impossible," and it "can be used to support any position or its opposite. If we build a nuclear power plant, it could melt down. If we don't build it, We will run short of power and society will collapse into anarchy" And worst, it "validates ignorance" because, "instead of focusing on what we know, it focuses on what we don't know-and what we can imagine." In the process, "risk assessment is devalued" and "probabilistic thinking is repudiated in favor of possibilistic thinking." As Schneier also notes, worst-case thinking is the driving force behind the precautionary principle, a decent working definition of which is "action should be taken to correct a problem as soon as there is evidence that harm may occur, not after the harm has already occurred." It could be seen in action less than a week after 9/11, when President George W Bush outlined his new national security strategy: "We cannot let our enemies strike first . . . [but must take] anticipatory action to defend ourselves, even if uncertainty remains as to the time and place of the enemy's attack. To forestall or prevent such hostile acts by our adversaries, the United States, will, if necessary act preemptively \_ . . America will act against such emerging threats before they are fully formed." The 2003 invasion of Iraq, then, was justified by invoking the precautionary principle based on the worst-case scenario in which Saddam Hussein might strike. If, on the other hand, any worst-case thinking focused on the potential for the destabilizing effects a war would have on Iraq and the region, the precautionary principle would guide one to be very cautious about embarking on war. As Sunstein notes, the precautionary principle "offers no guidance-not that it is wrong, but that it forbids all courses of action, including regulation." Thus, "taken seriously it is paralyzing, banning the very steps that it simultaneously requires."9 It can be invoked in equal measure to act or not to act. There are considerable dangers in applying the precautionary principle to terrorism: on the one hand, any action taken to reduce a presumed risk always poses the introduction of countervailing risks, while on the other, larger, expensive counterterrorism efforts will come accompanied by high opportunity costs." Moreover "For public officials no less than the rest of us, the probability of harm matters a great deal, and it is foolish to attend exclusively to the worst case scenario." A more rational approach to worst-case thinking is to establish the likelihood of gains and losses from various courses of action, including staying the current course." This, of course, is the essence of risk assessment. What is necessary is due consideration to the spectrum of threats, not simply the worst one imaginable, in order to properly understand, and coherently deal with, the risks to people, institutions, and the economy The relevant decision makers are professionals, and it is not unreasonable to suggest that they should do so seriously. Notwithstanding political pressures (to be discussed more in chapter 9), the fact that the public has difficulties with probabilities when emotions are involved does not relieve those in charge of the requirement, even the duty to make decisions about the expenditures of vast quantities of public monies in a responsible manner. [page 14-17]

#### 2] Predictions of rare events like extinction is next to impossible – prefer empirically verified events to improbable predictive analytics.

#### a] The future is unpredictable – the best way to preserve future value is to do good things now

Karnofsky 14 - Executive Director of the Open Philanthropy Project degree in Social Studies from Harvard University (Holden Karnofsky, 7/3/14, “The Moral Value of the Far Future” <https://www.openphilanthropy.org/blog/moral-value-far-future>)

I broadly accept the idea that the bulk of our impact may come from effects on future generations, and this view causes me to be more interested in scientific research funding, global catastrophic risk mitigation, and other causes outside of aid to the developing-world poor. (If not for this view, I would likely favor the latter and would likely be far more interested in animal welfare as well.) However, I place only limited weight on the specific argument given by Nick Bostrom in Astronomical Waste - that the potential future population is so massive as to clearly (in a probabilistic framework) dwarf all present-day considerations. More I reject the idea that placing high value on the far future - no matter how high the value - makes it clear that one should focus on reducing the risks of catastrophes such as extreme climate change, pandemics, misuse of advanced artificial intelligence, etc. Even one who fully accepts the conclusions of “Astronomical Waste” has good reason to consider focusing on shorter-term, more tangible, higher-certainty opportunities to do good - including donating to GiveWell’s current top charities and reaping the associated flow-through effects. More I consider “global catastrophic risk reduction” to be a promising area for a philanthropist. As discussed previously, we are investigating this area actively. More Those interested in related materials may wish to look at two transcripts of recorded conversations I had on these topics: a conversation on flow-through effects with Carl Shulman, Robert Wiblin, Paul Christiano, and Nick Beckstead and a conversation on existential risk with Eliezer Yudkowsky and Luke Muehlhauser. The importance of the far future As discussed previously, I believe that the general state of the world has improved dramatically over the past several hundred years. It seems reasonable to state that the people who made contributions (large or small) to this improvement have made a major difference to the lives of people living today, and that when all future generations are taken into account, their impact on generations following them could easily dwarf their impact in their own time. I believe it is reasonable to expect this basic dynamic to continue, and I believe that there remains huge room for further improvement (possibly dwarfing the improvements we’ve seen to date). I place some probability on global upside possibilities including breakthrough technology, space colonization, and widespread improvements in interconnectedness, empathy and altruism. Even if these don’t pan out, there remains a great deal of room for further reduction in poverty and in other causes of suffering. In Astronomical Waste, Nick Bostrom makes a more extreme and more specific claim: that the number of human lives possible under space colonization is so great that the mere possibility of a hugely populated future, when considered in an “expected value” framework, dwarfs all other moral considerations. I see no obvious analytical flaw in this claim, and give it some weight. However, because the argument relies heavily on specific predictions about a distant future, seemingly (as far as I can tell) backed by little other than speculation, I do not consider it “robust,” and so I do not consider it rational to let it play an overwhelming role in my belief system and actions. (More on my epistemology and method for handling non-robust arguments containing massive quantities here.) In addition, if I did fully accept the reasoning of “Astronomical Waste” and evaluate all actions by their far future consequences, it isn’t clear what implications this would have. As discussed below, given our uncertainty about the specifics of the far future and our reasons to believe that doing good in the present day can have substantial impacts on the future as well, it seems possible that “seeing a large amount of value in future generations” and “seeing an overwhelming amount of value in future generations” lead to similar consequences for our actions. Catastrophic risk reduction vs. doing tangible good Many people have cited “Astronomical Waste” to me as evidence that the greatest opportunities for doing good are in the form of reducing the risks of catastrophes such as extreme climate change, pandemics, problematic developments related to artificial intelligence, etc. Indeed, “Astronomical Waste” seems to argue something like this: For standard utilitarians, priority number one, two, three and four should consequently be to reduce existential risk. The utilitarian imperative “Maximize expected aggregate utility!” can be simplified to the maxim “Minimize existential risk!”. I have always found this inference flawed, and in my recent discussion with Eliezer Yudkowsky and Luke Muehlhauser, it was argued to me that the “Astronomical Waste” essay never meant to make this inference in the first place. The author’s definition of existential risk includes anything that stops humanity far short of realizing its full potential - including, presumably, stagnation in economic and technological progress leading to a long-lived but limited civilization. Under that definition, “Minimize existential risk!” would seem to potentially include any contribution to general human empowerment. I have often been challenged to explain how one could possibly reconcile (a) caring a great deal about the far future with (b) donating to one of GiveWell’s top charities. My general response is that in the face of sufficient uncertainty about one’s options, and lack of conviction that there are good (in the sense of high expected value) opportunities to make an enormous difference, it is rational to try to make a smaller but robustly positive difference, whether or not one can trace a specific causal pathway from doing this small amount of good to making a large impact on the far future. A few brief arguments in support of this position: I believe that the track record of “taking robustly strong opportunities to do ‘something good’ ” is far better than the track record of “taking actions whose value is contingent on high-uncertainty arguments about where the highest utility lies, and/or arguments about what is likely to happen in the far future.” This is true even when one evaluates track record only in terms of seeming impact on the far future. The developments that seem most positive in retrospect - from large ones like the development of the steam engine to small ones like the many economic contributions that facilitated strong overall growth - seem to have been driven by the former approach, and I’m not aware of many examples in which the latter approach has yielded great benefits. I see some sense in which the world’s overall civilizational ecosystem seems to have done a better job optimizing for the far future than any of the world’s individual minds. It’s often the case that people acting on relatively short-term, tangible considerations (especially when they did so with creativity, integrity, transparency, consensuality, and pursuit of gain via value creation rather than value transfer) have done good in ways they themselves wouldn’t have been able to foresee. If this is correct, it seems to imply that one should be focused on “playing one’s role as well as possible” - on finding opportunities to “beat the broad market” (to do more good than people with similar goals would be able to) rather than pouring one’s resources into the areas that non-robust estimates have indicated as most important to the far future. The process of trying to accomplish tangible good can lead to a great deal of learning and unexpected positive developments, more so (in my view) than the process of putting resources into a low-feedback endeavor based on one’s current best-guess theory. In my conversation with Luke and Eliezer, the two of them hypothesized that the greatest positive benefit of supporting GiveWell’s top charities may have been to raise the profile, influence, and learning abilities of GiveWell. If this were true, I don’t believe it would be an inexplicable stroke of luck for donors to top charities; rather, it would be the sort of development (facilitating feedback loops that lead to learning, organizational development, growing influence, etc.) that is often associated with “doing something well” as opposed to “doing the most worthwhile thing poorly.” I see multiple reasons to believe that contributing to general human empowerment mitigates global catastrophic risks. I laid some of these out in a blog post and discussed them further in my conversation with Luke and Eliezer. For one who accepts these considerations, it seems to me that: It is not clear whether placing enormous value on the far future ought to change one’s actions from what they would be if one simply placed large value on the far future. In both cases, attempts to reduce global catastrophic risks and otherwise plan for far-off events must be weighed against attempts to do tangible good, and the question of which has more potential to shape the far future will often be a difficult one to answer. If one sees few robustly good opportunities to “make a huge difference to the far future,” the best approach to making a positive far-future difference may be “make a small but robustly positive difference to the present.” One ought to be interested in “unusual, outstanding opportunities to do good” even if they don’t have a clear connection to improving the far future.

#### b] The world is complex – linear predictions are incoherent

Glover 12 (7/21/12, Robert W. Glover is the CLAS Honors Preceptor in Political Science at the University of Maine. “Compatibility or Incommensurability: IR Theory and Complex Systems Analysis” <http://www.e-ir.info/2012/07/21/compatibility-or-incommensurability-ir-theory-and-complex-systems-analysis/#_ftn1>)

A recent New York Times op-ed, written by a professor of political science no less, lambasted the discipline for consistently failing to predict both international and domestic political outcomes. It boldly proclaimed “…[c]himps throwing darts at possible outcomes would have done almost as well as the experts.”[1] To add insult to injury, the article featured a picture of a primate armed with darts taking aim at circular boards marked with regions of the world, possible outcomes, and likelihoods. The upshot of the article was that political science simply shouldn’t be in the business of prediction. Stevens quotes Karl Popper, stating “[l]ong term prophecies can be derived from scientific conditional predictions only if they apply to systems which can be described as well-isolated, stationary, and recurrent. These systems are very rare in nature; and modern society is not one of them.”[2] Though such indictments of our intellectual enterprise may be painful to hear, they are also correct in many senses. Political science, and by extension international relations (IR), has had difficulties in predicting future events with any accuracy or specificity. The guiding principles of “traditional” or “mainstream” approaches to IR have generally held that there is observable order in world affairs, from which we can offer explanations and make predictions. It is the great hope of our discipline that “there is an external world of which we can have knowledge…” and the notion that IR is “grounded in lawlike regularities that allow the possibility of making claims about how the ‘international’ operates.”[3] Yet if this were the case, surely we’d be doing a better job at forecasting international outcomes. The invocation of Popper reminds us why our best laid plans have gone awry. Social systems, from the most basic to the most intricate, almost invariably involve the complex interface of many variables, opaque interaction effects, and elements of chance and human variability. As Jervis states, “…[t]he result is that systems often display non-linear relationships, outcomes cannot be understood by the adding together of units or their relations and many of the results of actions are unintended. Complexities can appear in even what would seem to be simple and deterministic situations.”[4] Beyond prediction, even our attempts at post-hoc explanation tend to rely upon reductionism. That is to say we reduce the irreducibly complex to pithy “cause and effect” relationships. The 2008 war between Russia and Georgia was caused by a dispute over South Ossetia. The 1997 economic crisis was triggered by currency instability in Southeast Asia. The recent political upheaval in Egypt stemmed from technologically savvy young people angered with the corruption of Mubarak’s regime and a stagnant economy. These are the types of concise explanations we offer for events of enormous, systemic, complexity. However, IR theory has been grappling with a new set of tools which originate in the study of the natural world, specifically physics and biology. We call these tools “complex systems analysis” or in its more conceptual variant, “complexity theory.”[5] Complexity is not a unified theory as such, but rather an “emerging approach or framework” drawn from a variety of sources.[6] Proponents argue that IR can achieve better understanding of the world utilizing conceptual lenses attuned to the interaction of large numbers of variables and actors, interacting in a non-linear (and hence, less predictable) fashion. The remainder of this article will examine the rudiments of complexity theory, as well as its promise as a conceptual tool in understanding international relations. In particular, I will focus upon whether complexity theory constitutes a framework compatible with existing IR theories, or a fundamental and incommensurable challenge to the present theoretical landscape of IR

#### c] You can’t predict black swan events

Chadefaux 17 [Thomas Chadefaux (Department of Political Science, Trinity University); 20 February 2017; Data Science Journal; “Conflict forecasting and its limits”; <https://content.iospress.com/articles/data-science/ds002> //BWSWJ]

The question of predictability ultimately hinges on the underlying nature of conflict. In the words of Popper, does it more closely resemble the world of clouds – “highly irregular, disorderly, and more or less unpredictable” – or the one of clocks – “regular, orderly, and highly predictable in their behaviour” (Popper [37]). Unfortunately we do not yet know which we are facing. On most days, international and domestic interactions resemble a clock. Small deviations are corrected in a reversion to the mean, and the stochastic process of daily events and tensions that may emerge on a local or global level is trendstationary. Yet there are also rare shocks that do not follow this clock-like pattern. These are, of course, the events of interest here – conflicts, coups, acts of terrorism – that may start a cascade and change the clock into a cloud and the trend-stationary time series into unit-root processes (Doran [17]). There is yet a third possibility: that conflicts are neither clouds nor clocks, but black swans (Taleb [53]). Black swans are game-changing events with such low probability that they cannot be predicted (even though experts often claim to have found obvious warning signs for them ex post). Black swans are different from simple rare events. While rare events occur infrequently, their probability is not low conditional on the relevant set of variables. On the other hand, black swans have a low probability even conditional on other variables. Where conflict processes should be located on the clock-cloud-swan continuum matters. With clocks, predictions are possible, whether they be point or probabilistic predictions.7 With clouds, the marginal cost of better predictions would be increasing, but we could at least learn about the aggregate distribution and data-generating process (e.g. Clauset, Young, and Gleditsch [16] on the frequency of terrorist events). With black swans, however, attempting to predict would be a fool’s game. Several factors make it particularly challenging to predict conflict, and in fact may impose insuperable limits to our forecasts. First, our data are, almost by definition, prone to error and imprecision (e.g. Shellman, Hatfield, and Mills [47]). Part of it is due to poor measuring. But part of it is caused by strategic misrepresentation and concealing on the part of the relevant governments. A second reason for the difficulty to predict conflicts is that their structure, and more generally the structure of international relations, are constantly evolving. The end of the cold war, for example, was largely unanticipated and challenged many of the structures and patterns that formed parts of the existing models. Even within a given conflict, the dynamic can dramatically change and necessitate a different model (e.g. the surge in Iraq, Bhavnani et al. [4]). The difficulty is that these changes are difficult to anticipate – they are often black swans themselves – such that our ability to forecast may be limited to the short-term. The long term, on the other hand, would be the result of too many compounding shocks to a point where predictions become futile. Ideally, our predictions would be able to accomodate these changes. This may require two-level predictions in which the structure itself is first predicted, and within that structure the short-term events would be forecasted with a different model. But of course this would compound the uncertainty about our model specification and data. Another difficulty relates to the strategic nature of international relations and politics in general. First, actors are forward looking. They form their own predictions about the future, and act accordingly today. As a result, these predictions can affect their behavior today and invalidate these original predictions. As observers, then, we may have the right logic but end up not observing the phenomenon. Consider for example the problem of the onset of wars. If their contemporaries identify the underlying conditions as ripe for war, they may take additional steps to either prevent it, postpone its onset, or on the contrary speed it up, such that the initial predictions will be invalidated (e.g. Chadefaux [13] for empirical evidence of this pattern). The same logic also applies to the conduct and termination of war. The anticipations of forthcoming peace negotiations, for example, may lead certain actors – spoilers – to try harder to disrupt the peace process, thus reducing the prospects for peace (Kydd and Walter [31]). Mixed strategies are another difficulty. States or domestic actors cannot always respond to the same situation in the same way, else their response becomes predictable and may be exploited by the adversary. Just like a tennisman will not always serve in the same place to prevent the opponent from anticipating his actions, leaders must vary their threats and responses to events. As a result, the same conditions and sequence of events may lead to different reactions, some potentially leading to wars whereas others do not. In such cases, probabilistic predictions remain possible, but point predictions are inherently impossible. While a large number of observations may bring us close to an estimate of the underlying probability distribution, the predictive value of our forecasts will be bound upward by a fundamental limit. In fact, uncertainty itself may be necessary for the onset and continuation of war. Indeed, one of the central rationalist explanations for why bargaining might break down into war is incomplete information of at least one of the participants. As a result, “we cannot predict in individual cases whether states will go to war, because war is typically the consequence of variables that are unobservable ex ante, both to us as researchers and to the participants” (Gartzke [19], p. 567). Conflicts and the processes leading to them may also be path-dependent. A small event may lead to a cascading effect and ultimately to war. Yet the same underlying structure could possibly have accomodated an alternate equilibrium in which peace prevailed. Self-reinforcing processes mean that international interactions may magnify the effects of chance. Looking back, we may be able to trace the explosion of a keg – conflict – to a single spark (though this is itself debatable, as evidenced by the unfaltering scholarship on the causes of WWI), but looking forward, we are unable to know which spark will ignite the keg. In the same way, seismologists understand the causes of earthquakes and are able to monitor seismological variations with high precision but still cannot predict their onset with much early warning.8

#### 3] Extinction is inevitable – that non uqs magnitude weighing and means we ought to prefer more probable impacts that happen sooner

Monzon 20 Inigo Monzon [International Business TImes], 1-20-2020, "Mass Extinction After Asteroid Strike On Earth Is Inevitable, Scientists Reveal," International Business Times, <https://www.ibtimes.com/mass-extinction-after-asteroid-strike-earth-inevitable-scientists-reveal-2905584> DD AG

Scientists discussed the importance of studying past impact events to understand what asteroid strikes can do to Earth. According to one of the scientists, a major asteroid impact is inevitable and once it happens, it could wipe out almost all life on the planet.

The grim scenario was discussed in Discovery Channel’s series “Strip the Cosmos.” In the program, scientists from different fields discussed the impact asteroid strikes have on Earth.

Scientist Peter Schultz of the Ames Laboratory in California explained that studying previous impact events can provide valuable information regarding the effects of an asteroid strike on Earth. Through scientific investigations, Schultz learned that the asteroid that hit 66 million years ago in an area that is now known as Chicxulub, Mexico was not the first of its kind.

“Asteroids have changed the Earth, they have changed the life on Earth, the Chicxulub impact was not that rare of an event,” he said in the program according to Express.

As noted by other scientists, several impact events have already occurred on Earth over the course of millions of years. Some of these events were big enough to form large craters and even cause extinction events. Despite their size, many of these craters aren’t visible anymore due to the effects of Earth’s weather and geological activities.

For Laura Danly, the curator at Griffith Observatory in Los Angeles, Earth’s history with asteroid impacts indicate that a similar catastrophic event will happen again. Although determining exactly when an impact event will happen is a bit complicated due to a number of factors, Danly noted that a major asteroid strike is inevitable.

Once this happens, Danly said the Earth might experience the same disastrous events triggered by the asteroid impact 66 million years ago. In other words, a collision between the planet and a massive asteroid would trigger extreme environmental effects that would kill off more than half of all life on Earth.

“There’s no question that at some point in the future, an asteroid will strike the Earth with enough energy to extinguish most life, if not all,” she said.

#### 4] Actor specificity –

#### a] it is the moral obligation of the United States government to help its citizens, especially those who suffer the most from disease and poverty. There is no unique obligation for the US to prevent extinction that may or may not occur thousands of years in the future – fill in by other countries solves that and proves structural violence has to come prior

#### b] States can’t focus on abstract, overarching theories but should do what’s right because it’s the right thing to do – this commits them to consequentialism instead of inflexible rules

**Raz:** Raz, Joseph [Faculty, Columbia Law School] “Multiculturalism: A Liberal Perspective.” *Multiculturalism.* Winter 1994. RP //recut DD AG

**Political philosophy does not provide us with eternally valid theories for the government of all human societies. To my mind political philosophy is time-bound. It is valid—if it is valid at all—for the conditions prevailing here and now.** Its conclusions apply also to similar situations elsewhere. But we cannot set the precise boundaries for their application. There are two reasons for this limitation. First, **it is impossible to articulate comprehensively all the relevant moral considerations we are aware of, and impossible to state in general how much they weigh against each other in situations of conflict. Moral knowledge is practical in a special sense: it is embodied in our practices and acquired by habituation**. We often know what to do when faced with the situation in which action is called for when we could not have known what to do ahead of time. Everything we know can be articulated, can be ex- pressed in words. But it cannot be exhaustively expressed in general abstract formulae. The situation is analogous with that of a person who embarks on a journey to a distant destination. Ask him ahead of time to describe the route and he will be unable to do so. Yet as he progresses along the road he recalls at every stage how to proceed at that point. **Not everything we know can we exhaustively state in the abstract. Moral knowledge escapes such formulation, and that means that moral theories are to be taken as mere approximations. Those who apply them inflexibly are fanatics heading for disaster.**

#### 5] Default to probability – any other model of risk calculus collapses in on itself

Kessler 08 (Oliver; April 2008; PhD in IR, professor of sociology at the University of Bielefeld, and 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)

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 exempt**ed **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.

#### 6] Reject “1% risk of extinction”– it collapses all policymaking

**Meskill 09** (David, professor at Colorado School of Mines and PhD from Harvard, “The "One Percent Doctrine" and Environmental Faith,” Dec 9, http://davidmeskill.blogspot.com/2009/12/one-percent-doctrine-and-environmental.html)

Tom Friedman's piece today in the Times on the environment (http://www.nytimes.com/2009/12/09/opinion/09friedman.html?\_r=1) is one of the flimsiest pieces by a major columnist that I can remember ever reading. He applies Cheney's "one percent doctrine" (which is similar to the environmentalists' "precautionary principle") to the risk of environmental armageddon. But this doctrine is both intellectually incoherent and practically irrelevant. It is intellectually incoherent because it cannot be applied consistently in a world with many potential disaster scenarios. In addition to the global-warming risk, there's also the asteroid-hitting-the-earth risk, the terrorists-with-nuclear-weapons risk (Cheney's original scenario), the super-duper-pandemic risk, etc. Since each of these risks, on the "one percent doctrine," would deserve all of our attention, we cannot address all of them simultaneously. That is, even within the one-percent mentality, we'd have to begin prioritizing, making choices and trade-offs. But why then should we only make these trade-offs between responses to disaster scenarios? Why not also choose between them and other, much more cotidien, things we value? Why treat the unlikely but cataclysmic event as somehow fundamentally different, something that cannot be integrated into all the other calculations we make? And in fact, this is how we behave all the time. We get into our cars in order to buy a cup of coffee, even though there's some chance we will be killed on the way to the coffee shop. We are constantly risking death, if slightly, in order to pursue the things we value. Any creature that adopted the "precautionary principle" would sit at home - no, not even there, since there is some chance the building might collapse. That creature would neither be able to act, nor not act, since it would nowhere discover perfect safety. Friedman's approach reminds me somehow of Pascal's wager - quasi-religious faith masquerading as rational deliberation (as Hans Albert has pointed out, Pascal's wager itself doesn't add up: there may be a God, in fact, but it may turn out that He dislikes, and even damns, people who believe in him because they've calculated it's in their best interest to do so). As my friend James points out, it's striking how descriptions of the environmental risk always describe the situation as if it were five to midnight. It must be near midnight, since otherwise there would be no need to act. But it can never be five \*past\* midnight, since then acting would be pointless and we might as well party like it was 2099. Many religious movements - for example the early Jesus movement - have exhibited precisely this combination of traits: the looming apocalypse, with the time (just barely) to take action. None of this is to deny - at least this is my current sense - that human action is contributing to global warming. But what our response to this news should be is another matter entirely

#### **7]** Decision making should be cluster-based not sequence

Holden Karnofsky Executive Director of the Open Philanthropy Project degree in Social Studies from Harvard University; July 25, 2016; “Sequence thinking vs. cluster thinking” //BWSWJ // recut DD AG

When trying to compare two very different options (such as vaccinations and space colonization), it seems at first glance as though sequence thinking is superior, precisely because it allows huge numbers to carry huge weight. The practice of limiting the weight of uncertain perspectives can have strange-seeming results such as (depending on robustness considerations) giving equal weight to “Charity A seems like the better organization” and “Charity B’s goal is 200 billion times as important.” In addition, I find cluster thinking far more difficult to formalize and describe, which can further lower its appeal in public debates about where to give.

Below, I give several arguments for expecting cluster thinking to produce better decisions. It is important to note that I emphasize “better decisions” and not “correct beliefs”: it is often the case that one reaches a decision using cluster thinking without determining one’s beliefs about anything (other than what decision ought to be made). In the example given in the previous section, cluster thinking has not reached a defined conclusion on how likely space colonization is, how valuable space colonization would be, etc. and there are many possible combinations of these beliefs that could be consistent with its conclusion that supporting Charity A is superior. Cluster thinking often ends up placing high weight on “outside view” pattern-matching, and often leads to conclusions of the form “I think we should do X, but I can’t say exactly why, and some of the most likely positive outcomes of this action may be outcomes I haven’t explicitly thought of.”

The arguments I give below are, to some degree, made using different vocabularies and different styles. There is some conceptual overlap between the different arguments, and some of the arguments may be partly equivalent to each other. I have previously tried to use sequence-thinking-style arguments to defend something similar to cluster thinking (though there were shortcomings in the way I did so); here I use cluster-thinking-style arguments.

Sequence thinking is prone to reaching badly wrong conclusions based on a single missing, or poorly estimated, parameter

Sequence-style reasoning often involves a long chain of propositions that all need to be reasonable for the conclusion to hold. As an example, Robin Hanson lays out 10 propositions that cumulatively imply a decision to sign up for cryonics, and believes each to have probability 50-80%. However, if even a single one ought to have been assigned a much lower probability (e.g., 10^-5) – or if he’s simply failed to think of a missing condition that has low probability – the calculation is completely off.

In general, missing parameters and overestimated probabilities will lead to overestimating the likelihood that actions play out as hoped, and thus overestimating the desirability of deviating from “tried and true” behavior and behavior backed by outside views. Correcting for missed parameters and overestimated probabilities will be more likely to cause “regression to normality” (and to the predictions of other “outside views”) than the reverse.

Cluster thinking is more similar to empirically effective prediction methods

Sequence thinking presumes a particular framework for thinking about the consequences of one’s actions. It may incorporate many considerations, but all are translated into a single language, a single mental model, and in some sense a single “formula.” I believe this is at odds with how successful prediction systems operate, whether in finance, software, or domains such as political forecasting; such systems generally combine the predictions of multiple models in ways that purposefully avoid letting any one model (especially a low-certainty one) carry too much weight when it contradicts the others. On this point, I find Nate Silver’s discussion of his own system and the relationship to the work of Philip Tetlock (and the related concept of foxes vs. hedgehogs) germane:

Even though foxes, myself included, aren’t really a conformist lot, we get worried anytime our forecasts differ radically from those being produced by our competitors.

Quite a lot of evidence suggests that aggregate or group forecasts are more accurate than individual ones … “Foxes often manage to do inside their heads what you’d do with a whole group of hedgehogs,” Tetlock told me. What he means is that foxes have developed an ability to emulate this consensus process. Instead of asking question of a whole group of experts, they are constantly asking questions of themselves. Often this implies that they will aggregate different types of information together – as a group of people with different ideas about the world naturally would – instead of treating any one piece of evidence as though it is the Holy Grail. The Signal and the Noise, pg 66

In sequence thinking, a single large enough number can dominate the entire calculation. In consensus decision making, a person claiming radically larger significance for a particular piece of the picture would likely be dismissed rather than given special weight; in a quantitative prediction system, a component whose conclusion differed from others’ by a factor of 10^10 would be likely to be the result of a coding error, rather than a consideration that was actually 10^10 times as important as the others. This comes back to the points made by the above two sections: cluster thinking can be superior for its tendency to sandbox or down-weight, rather than linearly up-weight, the models with the most extreme and deviant conclusions.

A cluster-thinking-style “regression to normality” seems to prevent some obviously problematic behavior relating to knowably impaired judgment

One thought experiment that I think illustrates some of the advantages of cluster thinking, and especially cluster thinking that incorporates regression to normality, is imagining that one is clearly and knowably impaired at the moment (for example, drunk), and contemplating a chain of reasoning that suggests high expected value for some unusual and extreme action (such as jumping from a height). A similar case is that of a young child contemplating such a chain of reasoning. In both cases, it seems that the person in question should recognize their own elevated fallibility and take special precautions to avoid deviating from “normal” behavior, in a way that cluster thinking seems much more easily able to accommodate (by setting an absolute limit to the weight carried by an uncertain argument, such that regression to normality can override it no matter what its content) than sequence thinking (in which any “adjustments” are guessed at using the same fallible thought process).

The higher one’s opinion of one’s own rationality relative to other people, the less appropriate the above analogy becomes. But it can be easy to overestimate one’s own rationality relative to other people (particularly when one’s evidence comes from analyzing people’s statements rather than e.g. their success at achieving their goals), and some component of “If I’m contemplating a strange and potentially highly consequential action, I should be wary and seek robustness (not just magnitude) in my justification” seems appropriate for nearly everyone.

Sequence thinking seems to tend toward excessive comfort with “ends justify the means” type thinking

Various historical cases of violent fanaticism seem somewhat fairly modeled as sequence thinking gone awry: letting one’s decisions become dominated by a single overriding concern, which then justifies actions that strongly violate many other principles. (For example, justifying extremely damaging activities based on Marxist reasoning.) Cluster thinking is far from a complete defense against such things: the robustness of a perspective (e.g., a Marxist perspective) can itself be overestimated, and furthermore a “regression to normality” can encourage conformism with highly problematic beliefs. However, the basic structure of cluster thinking does set up more hurdles for arguments about “the ends” (large-magnitude but speculative down-the-line outcomes) to justify “the means” (actions whose consequences are nearer and clearer).

I believe that invoking “the ends justify the means” (justifying near and clear harms by pointing to their further-out effects) is sometimes the right thing to do, and is sometimes not. Specifically, I think that the worse the “means,” the more robust (and not just large in claimed magnitude) one’s case for “the ends” ought to be. Cluster thinking seems to accommodate this view more naturally than sequence thinking.

(Related piece by Phil Goetz: Reason as memetic immune disorder)

When uncertainty is high, “unknown unknowns” can dominate the impacts of our actions, and cluster thinking may be better suited to optimizing “unknown unknown” impacts.

Sequence thinking seems, by its nature, to rely on listing the possible outcomes of an action and evaluating the action according to its probability of achieving these outcomes. I find sequence thinking especially problematic when I specifically expect the unexpected, i.e., when I expect the outcome of an action to depend primarily on factors that haven’t occurred to me. And I believe that the sort of outside views that tend to get more weight in cluster thinking are often good predictors of “unknown unknowns.” For example, obeying common-sense morality (“ends don’t justify the means”) heuristics seems often to lead to unexpected good outcomes, and contradicting such morality seems often to lead to unexpected bad outcomes. As another example, expert opinion often seems a strong predictor of “which way the arguments I haven’t thought of yet will point.”