A Non-Credence account of the Long Term Future

I wrote a 20-page summary of *The Precipice* for my debate students to read and debate about. Most students thought that the normative claim that we should address existential risks is broadly correct, and that existential risks are a lot higher than they previously imagined. What they disagreed on the most was a descriptive thesis Ord presented: roughly, total existential risks within the next 100 years is 1 in 6 (pg 163)[1]. I'll refer to this thesis as P. I think this common ground is roughly analogous to what the average reader concludes after reading *The Precipice* : I learned new knowledge about how the world could plausibly end and about different existential risks, but 1 in 6 seems arbitrary. I withhold my belief on P (I am very uncertain over how likely it is for the world to end).

But it seems like Ord wanted to accomplish a lot more than that – that "this comparatively brief period is a unique challenge in the history of our species"(pg 40). In order for people to buy that we live in "The Precipice", then people must formulate some belief that there is a significant risk of existential risks.

If we are able to make readers hold stronger doxastic attitudes towards P, this would greatly help the overall persuasiveness of *The Precipice.* Then, I think what is missing from *The Precipice* is an even-if claim: Even if you do not believe in the credence attributed in P, why should you still believe that our society faces serious existential risks in the next 100 years. It is not a missing argument, but rather a missing technique on how to best persuade readers based upon an intuitive belief formation process.

In this essay, I aim to prove this missing technique in three ways. I'll first prove why a credence based account of P fails to be persuasive to the general public. I'll then illustrate several non-credence based accounts on how they make one hold a stronger doxastic attitude

[1] All citations with page numbers refer to The Precipice, Ord, Toby. *The precipice: Existential risk and the future of humanity*. New York: Hachette Books, 2021.

towards P; I'll lastly show the implications on persuasiveness if a non-credence based accounts were integrated alongside subjective probabilities in *The Precipice*.

1. Credence and Its Limitations on Persuasiveness

In Chapter 6 *Risk Landscape,* Ord uses credences in three ways:

Purpose	Example	Justification
(a) Represent ordinal ranking on the likelihood of existential risks from different events	X-risks from Unaligned Artificial Intelligence>Engineered Pandemics>Nuclear War/Climate Change (Pg 163)	Evaluation of impact magnitude/Whether we have sufficient safety factors to guard against these risks
(b) Verify probabilistic arguments with reality	Natural risks can't be that high, or else it is unlikely that we would have survived until now (pg 161)	Intuitive probabilistic claim that it is likely we won't be that lucky if natural risks were magnitudes higher
(c) Showcase a roughly correct credence about Existential Risks	Total extinction chance is 1 in 6, AI extinction chance is 1 in 10 (pg 163)	Best estimates given current information and argumentation (pg 162)

I think the average reader will probably be convinced by (1) and (2) but not (3). There are a list of reasons why the average reader would not find (3) to be persuasive.

1. People do not usually reason with subjective probabilities.[2] People tend to have flawed intuitions around subjective probabilities and what they actually entail.[3] People also tend to be naturally hesitant towards buying subjective probabilities about big claims, even if they are well reasoned.

[2] Jackson, Elizabeth G. "The Relationship between Belief and Credence." *Philosophy Compass* 15, no. 6 (2020). https://doi.org/10.1111/phc3.12668.

[3] "Measuring People's Preferences." IDinsight. Accessed June 16, 2023. https://www.idinsight.org/publication/measuring-peoples-preferences/.

2. A subjective probabilities account will be really hard for policy makers to defend under close scrutiny. This is because you have to offer externalist reasons why you chose this credence – but credences are naturally fuzzy.[4]

3. It is often very difficult to enunciate arguments justifying subjective probabilities[5]. There are often key information gaps and uncertainties that we are making best estimates for. We tend to be clueless about the long term future.

Most people are not Bayesian and won't be after reading *The Precipice.* I think Ord is right that it is useful to quantify beliefs despite uncertainties, but in addition to defending a credence based account, it is missing a way to convince people even if they do not believe in subjective probabilities presented by Ord.

2. **Non-Credence Based Account**

I think speculating about long-term existential risks could be aided by Information Gap Decision Theory (or IGDP). *Information Gap* is the disparity between what needs to be known and what is known. A *robustness strategy* satifices the outcome and guards against error or surprises.[6]

First, map out information gaps we have about existential risks. I think Ord does this well by marking his uncertainties in different chapters, particularly in Chapter 5 . Specifically what is missing here is to identify a series of events which will close information gaps we have about existential risks. These are early indicators which say: even if you do not believe in the conclusion of *The Precipice* now, what are some future events which will make you

---

[4] Lyon, Aidan. "Vague Credence." *Synthese* 194, no. 10 (2015): 3931–54. https://doi.org/10.1007/s11229-015-0782-5.

[5] Ross, Jacob, and Mark Schroeder. "Belief, Credence, and Pragmatic Encroachment." *Philosophy and Phenomenological Research* 88, no. 2 (2012): 259–88. https://doi.org/10.1111/j.1933-1592.2011.00552.x.

[6] Ben-Haim, Yakov. "Implications of Info-Gap Uncertainty." *Info-Gap Decision Theory*, 2006, 317–46. https://doi.org/10.1016/b978-012373552-2/50014-x.

believe in existential risks. This looks like if Artificial Intelligence exhibits power-seeking behaviour, or if there are examples of low-harm misaligned AIs. If *The Precipice* is correct, then this will be an even more important book in the future – identifying points on when readers will discover and change their beliefs makes the book more persuasive. It is finding a way for people to buy in Bayesian Evidence to update their beliefs.

Second, offer a leap-of-faith approach on why we should believe in existential risks being high even if there are information gaps. Ord hints at this, where he said "... to start with an extremely small probability and only raise it from there when a large amount of hard evidence is presented. But I disagree. Instead, I think the right method is to start with a probability that reflects our overall impressions, then adjust this in light of the scientific evidence". I think Ord needed to defend our overall impression approach a lot more, as I think this could be a really useful rhetorical tool. This would look like defending that we often formulate beliefs about something despite information gaps(on a day to day basis with epistemic leap of faith); Or how the former raising from small probabilities is an impossible task to do when we do not know about the long term future; Or how reasoning about the long term future requires you to assume some things as true, even if you don't have strong reasons for it.

Third, advocate for a *robustness strategy* even if we do not believe in P. Ord argued for P not because of the epistemic value for P, but that we should act in ways to significantly reduce existential risks. If we can reach the same terminal goal, then P is no longer needed as an instrumental goal. Drawing analogies here with existing policies would be persuasive to the general public and offer good rhetorical tools for supportive policy makers. For example, we have "no direct evidence of cell phones or other electronic devices interfering with aircraft systems" but yet we ban cell phones on planes simply because of how bad a crash

would be.[7] The same can be said about why mitigating existential risks (at some cost) would be fine without direct evidential justification for any particular credence – but simply because of how important the situation is.

3. Implications

I wrote about the non-credence based account in a fairly abstract way – and that is because I wish to lay out the theoretical foundations first on what people will find to be persuasive if they do not believe in subjective probabilities. Tractable implications on how The Precipice would look like 1.) ways to reduce how conclusions of *The Precipice* were linguistically dependent on subjective probabilities 2.) come up with more examples such as the Airplane-Signal one, and integrate them into *The Precipice* 3.) altering the appendix to include some of the even-ifs.

I don't think my argument changes the conclusion of *The Precipice*. It changes how the conclusion is presented. I think this is really important – because objections on increasing or decreasing the likelihood and harm of specific existential risks seems like a fine-tuning of *The Precipice* – but I think the book is pretty well-argued to begin with and it is likely that the changes won't be decision-relevant information for the reader.

I really do think we live in "The Precipice" of our time. It is best if more readers believe this as well.

---

[7] Hsu, Jeremy. "The Real Reason Cell Phone Use Is Banned on Airlines." LiveScience, December 21, 2009. https://www.livescience.com/5947-real-reason-cell-phone-banned-airlines.html.

Bibliography

Ben-Haim, Yakov. “Implications of Info-Gap Uncertainty.” *Info-Gap Decision Theory*, 2006, 317–46. https://doi.org/10.1016/b978-012373552-2/50014-x.

Hsu, Jeremy. “The Real Reason Cell Phone Use Is Banned on Airlines.” LiveScience, December 21, 2009. https://www.livescience.com/5947-real-reason-cell-phone-banned-airlines.html.

Jackson, Elizabeth G. “The Relationship between Belief and Credence.” *Philosophy Compass* 15, no. 6 (2020). https://doi.org/10.1111/phc3.12668.

Lyon, Aidan. “Vague Credence.” *Synthese* 194, no. 10 (2015): 3931–54. https://doi.org/10.1007/s11229-015-0782-5.

“Measuring People’s Preferences.” IDinsight. Accessed June 16, 2023. https://www.idinsight.org/publication/measuring-peoples-preferences/.

Ord, Toby. *The precipice: Existential risk and the future of humanity*. New York: Hachette Books, 2021.

Ross, Jacob, and Mark Schroeder. “Belief, Credence, and Pragmatic Encroachment.” *Philosophy and Phenomenological Research* 88, no. 2 (2012): 259–88. https://doi.org/10.1111/j.1933-1592.2011.00552.x.