In Defense of Timidity

There are many moral issues but limited resources to address them, we must carefully choose which one to prioritize. Effective Altruism (EA), a social and philosophical movement created at the turn of the 2010s, focuses on this prioritization challenge. EA attempts to identify the most effective ways to use our resources to maximize welfare and is committed to supporting the identified most effective initiatives (MacAskill, 2019, 14). Historically, EA has prioritized helping theory, the risky prospect is a thousand times better. However, its probability is so low that choosing it will likely result in 1000 people dying for nothing. The LCA is based on comparisons of similar prospects. The prospect with a tiny probability of a large payoff is similar to the long-term intervention, while the sure prospect is similar to the short-term intervention.

recklessness. More precisely, I argue that a *lexical discounting* view with a *range threshold* is more plausible than recklessness and, thus, that it is morally permissible for individuals not to prioritize long-term interventions. In other words, the moral duty entailed by the LCA can only apply to societies.

In part 1, I set the table of the issue and specify the timid view I defend, a discounting view with a range threshold. In part 2, I present the five most important objections to recklessness, objections that reckless people have, at least not yet, responded to. In part 3, I present the three most important objections to discounting views and show that the view I defend responds to two of them. I also present a mathematical solution to third objection.

In part 4, I summarize, conclude, and present the implications of my argument for longtermism.

1. Setting the Table

The issue I address in this paper requires some technical clarifications. In section 1.1, I present some background clarifications. In section 1.2, I present the paradox of recklessness to set the issue at hand. In section 1.3, I define the two competing views addressing the paradox: recklessness and timidity. In section 1.4, I present a novel discounting view using a range threshold instead of the standard fixed threshold. I ground this novelty in a contextualist view addressing the

Third, the theory of the value of uncertain prospects assumed in this paper is the standard expected value theory (EVT). According to EVT, an outcome is equal to the product of the value of its expected gain or loss by the probability that it will occur. The value of the prospect is the weighted sum of the outcomes. Moreover, according to EVT, a prospect is better than another if and only if it has a higher expected value. This is the most common normative model used to evaluate simple uncertain prospects and the model used by longtermists.

Third, recklessness and timidity can be positive or negative. The positive version addresses good prospects (e.g., people having a good life), and the negative version addresses bad prospects (e.g., lives of suffering). For the sake of brevity and simplicity, I only address the positive version, but the discussion can also apply to negative versions of the views with some adjustments.

Finally, the issue I address in this paper has recently received increased attention because of its importance for longtermism. Three influential publications were recently published on the topic by longtermists or longtermist-adjacents researchers (Beckstead & Thomas, 2023; Wilkinson, 2022; Kosonen, 2022). Wilkinson fully embraces recklessness and strongly rejects timidity. Beckstead and Thomas reject both recklessness and timidity. Kosonen also rejects both views, but she seems, *inter alia*, to consider discounting views to be the most plausible view.

(deal 0) is 0.999^0, i.e., 1; the probability of the devil's deal 1 is 0.999^1, i.e., 0.999; the probability of the second devil's deal is 0.999^2, i.e., 0.998, and so forth. The payoff is 10^n, nothing otherwise. Deal 0 has a payoff of 10^0

1.3. The Definitions of Recklessness and Timidity

Beckstead and Thomas (2023, 4) define recklessness as follows:

> ***Recklessness***: 1) For any finite payoff *x*, no matter how good, and any probability *p*, no matter how tiny, 2) there's a finite payoff *y*, such that getting *y* with probability *p* is better than 3) getting *x* for sure. (My numeration)

To make the definition more concrete, let me give an example with arbitrary numbers. For the approach seems implausible.

addressed in the extensive literature on the philosophy of language and the philosophy of sciences regarding how to deal with vagueness.[4]

The main views to address vagueness are the many-valued logic views, supervaluationism, subvaluationism, and contextualism.[5] As a rough overview, many-valued logic views hold that borderline cases have a truth value between full falsehood and full truth. Supervaluationism holds that borderline cases are neither true nor false. Subvaluationism holds that borderline cases are both true and false. Contextualism holds that the set of objects to which a term applies changes according to the context.

I hold the latter view but will assume it to be the most plausible for brevity. What contextualists refer to as being part of the context can refer to many aspects: who is talking or making a decision, what objects are in the set, external conditions, etc. I will emphasize the importance of accounting for the objects of the sets to clarify where to put the threshold for a particular situation.

For instance, Tom is six feet tall. Is he above the threshold for tallness? Being tall is vague and depends on contextual elements. For the general population, six feet is tall, but in the NBA, this is short. Being six feet tall is a borderline case that requires knowing the set of objects to set the threshold for tallness correctly. A way to see it is that we can use a range threshold for tallness:

The first step is to set a fixed threshold as a starting point for handling unambiguous cases. For instance, suppose the threshold is set at a probability of one-trillionth, and the range is between one-billionth and one-quadrillionth. If prospect A has a probability of one-millionth and prospect B of one-hundred-quadrillionth, both cases are outside of the range threshold. We should then simply use the fixed threshold.

The second step is to set the width of the threshold with a lower limit below which prospects would have to be discounted regardless of the context and an upper limit above which no prospect could be discounted regardless of the context.

The third step is to set up the rules to adjust the threshold for specific cases. Three rules for adjustments apply to the cases I address.

Rule 1: We can adjust the fixed threshold when the expected value of a prospect above the fixed threshold is comparatively *much smaller* than one of the prospects below the fixed threshold but above the range threshold. This avoids inconsistent ordering when huge payoffs are barely below the threshold compared to tiny payoffs.

Rule 2: We can make adjustments when two prospects of similar payoffs have just a tiny difference in probability, one being slightly above and the other slightly below the fixed threshold. This avoids differential treatment of otherwise similar outcomes.

Rule 3: We should discount prospects with negligible payoff to avoid a prospect with a high probability of getting virtually nothing being preferred to a tiny probability of something of impact.

The objections presented in part 3 are prime examples of when the range threshold ought to be used. Before testing the range threshold, let me justify the need to discount tiny probabilities by showing how a non-discounting approach, recklessness, ought to be discarded as a plausible view to deal with prospects with tiny probabilities.

Answers compatible with recklessness have been proposed. For instance, Feller argues that a casino must guarantee to have access to an infinite amount of money in case the gambler gets

Regarding the consequent, the principle then states that it should be true that prospect A (which has an infinite expected value) is strictly better than prospect B (which has an infinite expected value). However, both are equal, they have an infinite expected value. The consequent is false.

As a conditional with a true antecedent and a false consequent is false, the reckless interpretation of the St-Petersburg gamble violates Prospect-outcome dominance. If the expected value of the St-Petersburg gamble is not infinite, as is the case for timid interpretations, some outcomes of

2.4. Decision Swamping and Lack of Risk-Aversiveness

I will add to the list of objections to recklessness that it leads to extreme risk-seeking by having outcomes with infinite or huge payoffs swamping decisions, even if all the other outcomes are extremely harmful and overwhelmingly probable. In other words, recklessness is not enough risk-averse.

Let us imagine two prospects, A and B. Prospect A has two potential outcomes, one with an estimated probability *p* of 10^{-30}

(that might never occur), hundreds of trillions of lives have to be sacrificed. This is especially troubling if we think of our epistemic conditions. For instance, how do we know that the project will not be aborted before we average out the losses? Millions of lives would have been wasted to get no gain. However, there is a lot of uncertainty, leaving room for a reasonable reply by longtermists to this objection.

However, for the Devil's deals thought experiment, we clearly see the irrationality of recklessness. There is a probability of 10^{-21} to get the last Devil's deal, the preferred option by recklessness. Respecting the conditions that the law of large numbers entails for the last deal to be worth the shot would require sextillions of trials to average out the losses. However, recklessness advises going for the last deal with *a single trial*! This is, in my opinion, the worst objection to recklessness, it is insensitive to the number of trials, an essential aspect to account for.

The law of large numbers best explains the paradox of recklessness. The first deals meet the law of large numbers. If one has a probability of 99.7% of getting 1000 years of life, a single trial is sufficient to average the losses. For the first hundreds of deals, a single trial is sufficient, and recklessness is thus a good guide, as is timidity. However, as the probability gets lower, more and more trials are necessary to average the losses and have the expected value not to count on luck. With a probability around 0.1, a couple of trials are more than enough, and for the last deals, sextillions of trials would be necessary. The problem is that the Devil's deal precludes meeting this condition, a single ticket has to be chosen. By contrast, discounting views avoid this issue by discounting the deals that do not meet the necessary conditions to average out the losses. The number of trials is a contextual aspect that the contextualist view I hold accounts for.

3. Objections to Discounting views

As we saw, recklessness entails some very disturbing conclusions. To avoid this conclusion, we have to bind the function in some ways, as all the problems are derived from the function not being bounded. Discounting tiny probabilities is the most promising approach to binding functions.

Second, the worst issue is that a prospect with an insignificant payoff just above the threshold

One might think that we could change the values to make the objection still relevant to account for my solution. However, no modification of the values of the objection would save it. The most difficult to answer modification is the comparison between P_2 and P_3. The two modifications with the most potential are to reduce the value of probability *q*, to make it outside of the range threshold so no adjustment would make P_2

Table 4: Payoffs for Different Prospects: Original Background Independence

fective modifications to attack my view would make the distribution of the probability between *p* and *q* different, making the values either closer or farther apart. On the one hand, if one makes *p* larger and *q* smaller, it would make the adjustment of lowering the threshold easier as we would remove a smaller *q* for the first strategy I used. On the other hand, if we make *p* smaller and *q* larger, this changes nothing for the second strategy, as we just have to make *p*+*q* below the threshold slightly higher by a micro adjustment to discount all prospects and use the tail discounting. Because we can go both ways, making the threshold slightly lower or higher, no modification of the values of *p* and *q* saves the objection.

2.3. Objection 3: Continuity Axiom

The last objection to discounting views is presented by (Kosonen, 2022, 200-201) and involves the violation of the Continuity axiom. Let us assume that for all prospects X and Z, XpZ represents the prospect of a lottery where X has a probability *p* of realizing and Z has a probability 1

Response

This violation only occurs with the use of a clear-cut discount. It can be easily avoided by using a continuous discount function.

A clear-cut discount may be formalized mathematically by multiplying the utility function of a prospect X by a discontinuous « step-function », widely called the Heaviside function:

Equation (1)

We then obtain a discounted prospect X_d, whose utility is a function of probability p:

Equation (2)

As a result, Xd=0 on the domain p (0,t) and Xd=X for p

Figure 2: Continuous functions approaching the Heaviside step function, for increasingly narrow step width

y , probability threshold " , step function defined within the range " . In red: the Heaviside step function.

As the discount function can be adjusted to be arbitrarily close to its extremely discontinuous version, the "width of the step" can be arbitrarily narrowed, for example, to fit between two prospects. Then, in the context of calculations not involving this sensitive zone, the results will be the same as for using a perfect step function, which can then be used for simplicity's sake. Note that it can be shown that the continuity axiom would also be preserved if using any other continuous discount functions such as an exponential discount.

4. Conclusion

In conclusion, I start by summarizing the objections to timidity and its answers and then summarize the objections to recklessness. After this summary, I present a concluding analysis before presenting the implications of my argument for longtermism.

In part 2, we saw that recklessness faces at least five objections. First, recklessness entails paying any finite amount of money to participate in St-Petersburg-type gambles, an unreasonable conclusion. The St-Petersburg gamble has nothing special about it, so the issue that recklessness faces generalizes to a large variety of decision situations. Second, recklessness violates Prospect-

ibility to address such challenges. Anyhow, in the current state of research, recklessness seems to obviously be a worse view than a refined discounting view.

As stated in the introduction, the longtermist comparative argument supporting the moral obligation of individuals to support long-term interventions depends on recklessness. However, as we have seen in this paper, recklessness is less plausible than the discounting view I defended. The argument on which Greaves and MacAskill base strong longtermism is thus in jeopardy for individuals. The debate on recklessness leads to the intuitive conclusion that societal projects like protection against existential risks are a duty for societies and that it should be morally permissible for individuals to support other existing individuals effectively with their donations or by choosing careers based on their preference rather than having the duty to maximize good for the future. In sum, even consequentialist views do not entail longtermism for individuals.

MacAskill, W. (2015). *Doing Good Better: Effective Altruism and a Radical New Way to Make a Difference.* Random House.

McMahan, J. (1981). ‘Problems of population theory’. *Ethics*, vol. 92, no. 1, 96–127. https://doi.org/10.1086/292301

Mehlberg, H. (1958). The Reach of Science. Toronto. Toronto University Press.

Mogensen, A. L. (2022). ‘The only ethical argument for positive d?’. *Philosophical Studies*, vol. 179, 2731–50.

Ord, T. (2020). The Precipice: Existential Risk and the Future of Humanity. Hatchette, 480 pages.

Peterson, M. (2002). ‘What is a de Minimis risk?’. *Risk Management*, vol. 4, 47–55.

Raffman, D. (1996) ‘Vagueness and context-sensitivity’. *Philosophical Studies*, vol. 81, 175–92.

Russel, J. S. (2023). On two arguments for Fanaticism. Noûs, online version before inclusion in an issue.

Samuelson, P. (1960). 'The St. Petersburg Paradox as a Divergent Double Limit’. *International Economic Review*, vol. 1, no. 1, 31–7.

Shapiro, S. (2006). *Vagueness in Context.* Oxford: Oxford University Press.

Singer, P. (2015). The Most Good You Can Do: How Effective Altruism is Changing Ideas About Living Ethically. Yale University Press.

Singer, P. (1972). Famine, affluence, and morality. *Philosophy and Public Affairs*, vol. 1, 229–243.

Smith, N. J.J. (2014). Is evaluative compositionality a requirement of rationality? *Mind*, 457–502.

Smith, N. J. J. (2008). *Vagueness and Degrees of Truth*, New York: Oxford University Press.

Tarsney, C. (2023). The epistemic challenge to longtermism. *Synthese, vol.* 201, no. 6, 1-37

Temkin, L. (2015). Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning: A Précis.