Pascal's Hell*

Petra Kosonen[†]

January, 2024

ABSTRACT: Fanaticism is the idea that tiny probabilities of huge positive or negative payoffs can have enormous positive or negative expected value (respectively). Fanaticism is counterintuitive, as shown by cases such as Pascal's Mugging and the St. Petersburg paradox. However, these cases do not illustrate the most counterintuitive implications of fanaticism. This paper shows, in the form of a dialogue, how a world filled with fanatical agents can be almost certain to be void of value—and by their own choice.

Satan: I have an offer for you, Pascal, as I have heard that you might be interested in a small probability of a huge payoff.

Pascal: Anything that maximizes expected utility!

*I wish to thank Johan Gustafsson, Samuel Marks, Andreas Mogensen and Teru Thomas for valuable comments.

[†] Population Wellbeing Initiative, University of Texas at Austin. I would be grateful for comments: kosonenpetra@gmail.com Satan: Great. And your utility function is unbounded, am I right?¹

Pascal: Yes, and additive in terms of people's happy days of life.

Satan: Excellent. I believe you have already met my dominion, the Mugger.² And you refused his offer. Would you remind me of the reasons for your refusal? *Pascal*: I had a few reasons for refusing the offer. First, the larger the payoff the Mugger promised to deliver, the lower the probability I assigned to him sticking with his promise.³ And my probabilities went down faster than the Mugger's offer's utilities went up. Secondly, there exist symmetrical possibilities that cancel out the expected value of the Mugger's offer.⁴ For example, I might meet another mugger who promises me an amazing payoff if I do not pay the original mugger.

Satan: Very well. And you also refused to pay a large sum to play the St. Petersburg game, is that correct?⁵

¹The utility function does not necessarily have to be unbounded for this case to work—it is enough that the upper bound is very high and the lower bound very low.

²In Bostrom's (2009) Pascal's Mugging, a mugger claims to be an Operator from the Seventh Dimension and promises to give Pascal a thousand quadrillion happy days in the Seventh Dimension if Pascal pays the mugger ten livres—money that the mugger will use for helping many orphans in the Seventh Dimension. This case is based on discussions by various people, such as Yudkowsky (2007).

³This objection is from Baumann (2009, p. 447). See Hanson (2007) for a related objection.

⁴This objection is from Hiller and Hasan (forthcoming, p. 14).

⁵The St. Petersburg game, originally proposed by Nicolaus Bernoulli in 1713, goes as follows: A fair coin is flipped until it lands on heads. The prize is 2^n , where *n* is the number of coin flips. See Pulskamp (2013). *Pascal*: Yes, I'm suspicious of infinities and do not know how to handle them.⁶ *Satan*: Fair enough. I have a new offer for you, and it does not involve infinities.⁷ Also, your concerns about the Mugger's offer don't apply here either. So, the offer is this: I will flip a coin, and if it lands on heads, I will help humanity settle on new planets in faraway galaxies and live in bliss until the heat death of the Universe. (This is the fate of the Universe in which Pascal and Satan live.) Until the heat death happens, it will be like heaven. But if the coin lands on tails, then everyone on Earth will suffer excruciating pain until life on Earth is no longer possible—one billion years of hell. If you don't accept the offer, humanity will live as mere mortals until life on this planet is no longer possible. Humanity will not be able to expand out from Earth without my help.

Pascal: That sounds great—even odds of utopia!

Satan: Oh, pardon me, I forgot to say that my coin is somewhat biased. The probability of heads happening is one-in-a-googolplex. I admit the odds aren't great. But if the coin lands on heads, I will create a thousand googolplex happy Earth-like planets.

Pascal: Not to worry, the offer is still amazing. The expected value of taking the

⁶Russell and Isaacs (2021), Russell (forthcoming) and Beckstead and Thomas (forthcoming).

⁷Pascal's Mugging is similar to Pascal's Wager, except that the former does not involve infinite utilities. Pascal (1958) famously argued that one should believe in God because of the possibility of gaining an infinitely good payoff in Heaven: "Let us weigh the gain and the loss in wagering that God is. Let us estimate these two chances. If you gain, you gain all; if you lose, you lose nothing. Wager, then, without hesitation that He is." gamble is clearly greater than the expected value of rejecting it. Actually, its expected value might even be greater than the expected value of the offer I initially thought you were making... So, I'm positively surprised.

Satan: Oops, I made a mistake: I read the wrong page. The instruction manual (*Creating Hell*) says that the probability of heads happening on Earth is one-in-Graham's-number. But it is in my power to create any finite number of happy Earth-like planets, so I believe I can still give you a good offer. If the coin lands on heads, I will create a million Graham's number of happy Earth-like planets.

Pascal: Now your offer is even better! Although I dread the almost certain torture for everyone on Earth for the next billion years, the expected value of your offer is far greater than the expected value of not taking it. So, rationality compels me to accept it.

Satan: I never imagined persuading people to enter (finite) hell would be this easy...

* * *

So Satan traveled from one planet to another, and the inhabitants of those planets also expected utility maximizers with unbounded utilities—always accepted his offer. According to Satan's instruction manual, the probability of the coin ever landing on heads was merely one-in-a-googolplex, so the Universe was almost certain to be void of joy and laughter. Yet they all lived happily ever after (in expectation).

References

Baumann, P. (2009), 'Counting on numbers', Analysis 69(3), 446-448.

- Beckstead, N. and Thomas, T. (forthcoming), 'A paradox for tiny probabilities and enormous values', *Nous*.
- Bostrom, N. (2009), 'Pascal's Mugging', Analysis 69(3), 443-445.
- Hanson, R. (2007), 'Pascal's Mugging: Tiny probabilities of vast utilities'.
 URL: https://www.lesswrong.com/posts/a5JAiTdytou3Jg749/pascal-s-muggingtiny-probabilities-of-vast-utilities?commentId=Q4ACkdYFEThA6EE9P
- Hiller, A. and Hasan, A. (forthcoming), 'How to save pascal (and ourselves) from the mugger', *Dialogue*.
- Pascal, B. (1958), *Pascal's Pensées*, E. P. Dutton & Co., New York. URL: https://www.gutenberg.org/files/18269/18269-0.txt
- Pulskamp, R. J. (2013), 'Correspondence of Nicolas Bernoulli concerning the St. Petersburg Game'. Unpublished manuscript. Accessed through: https://web.archive.org/.
 - **URL:** http://cerebro.xu.edu/math/Sources/NBernoulli/correspondence_petersburg_game.pdf
- Russell, J. S. (forthcoming), 'On two arguments for fanaticism', Nous .
- Russell, J. S. and Isaacs, Y. (2021), 'Infinite prospects', *Philosophy and Phenomeno-logical Research* **103**(1), 178–198.

Yudkowsky, E. (2007), 'Pascal's Mugging: Tiny probabilities of vast utilities'. **URL:** *http://www.overcomingbias.com/2007/10/pascals-mugging.html*