Pascal's mugging

NICK BOSTROM

In some dark alley...

Mugger: Hey, give me your wallet.

Pascal: Why on Earth would I want to do that?

Mugger: Otherwise I'll shoot you.

Pascal: But you don't have a gun.

Mugger: Oops! I knew I had forgotten *something*.

Pascal: No wallet for you then. Have a nice evening.

Mugger: Wait!

Pascal: Sigh.

Mugger: I've got a business proposition for you.... How about you give me your wallet now? In return, I promise to come to your house tomorrow and give you *double* the value of what's in the wallet. Not bad, eh? A 200% return on investment in 24 hours.

Pascal: No way.

Mugger: Ah, you don't believe that I will be as good as my word? One can't be too careful these days.... Tell you what: give me your wallet, and I come to your house tomorrow and pay you *10 times* its value.

Pascal: Sorry.

Mugger: OK, let me ask you something. Many people are dishonest, but some people are honest. What probability do you give to the hypothesis that I will keep my promise?

Pascal: 1 in a 1,000?

Mugger: Great! OK, so give me your wallet, and tomorrow I give you *2,000 times* the value of its contents. The expectation value is greatly to your advantage.

Pascal: There are 10 livres in my wallet. If we made a deal for you to take the wallet and bring me 10 times the value of its contents tomorrow, then maybe there's a 1-in-a-1,000 chance that I would see the 100 livres you owe. But I'd rate the chances that you will deliver on a deal to return me 20,000 livres much lower. I doubt you even have that much money.

Mugger: Your scepticism is understandable, although in this particular case it happens to be misguided. For you are M. Pascal if I'm altogether not mistaken? And I've heard that you're a committed expected-Utility maximizer, and that your Utility function is aggregative in terms of happy days of life. Is that not so?

Analysis Vol 69 | Number 3 | July 2009 | pp. 443–445 doi:10.1093/analys/anp062

Pascal: It is. My Utility function is unbounded. And I deem two days of happy life twice as good as one such day; and 2,000 days twice as good as 1,000 days. I don't believe in risk aversion or temporal discounting.

Mugger: Excellent. I don't necessarily have to know that you reject risk aversion and temporal discounting, but it makes things easier. Well, have I got good news for you! I have magical powers. I can give you any finite amount of money that you might ask for tonight. What's more, I can give you any finite amount of Utility that I choose to promise you tonight.

Pascal: And I should believe you why?

Mugger: Trust me! OK, I realize this does not give you conclusive evidence, but surely it counts a least *a little bit* in favour of the truth of what I am asserting. Honestly, I really do have these powers.

Pascal: Your conduct tonight has not inspired me with confidence in your honesty.

Mugger: OK, OK, OK, OK. But isn't it *possible* that I am telling the truth?

Pascal: It is *possible* that you have the magic powers that you claim to have, but let me tell you, I give that *a very, very low probability*.

Mugger: That's fine. But tell me, how low a probability exactly? Remember, you might think it all seems implausible, but we are all fallible, right? And you must admit, from what you've already seen and heard, that I am a rather atypical mugger. And look at my pale countenance, my dark eyes; and note that I'm dressed in black from top to toe. These are some of the telltale signs of an Operator of the Seventh Dimension. That's where I come from and that's where the magic work gets done.

Pascal: Gee . . . OK, don't take this personally, but my credence that you have these magic powers whereof you speak is about one in a quadrillion.

Mugger: Wow, you are pretty confident in your own ability to tell a liar from an honest man! But no matter. Let me also ask you, what's your probability that I not only have magic powers but that I will also use them to deliver on any promise – however extravagantly generous it may seem – that I might make to you tonight?

Pascal: Well, if you really were an Operator from the Seventh Dimension as you assert, then I suppose it's not such a stretch to suppose that you might also be right in this additional claim. So, I'd say one in 10 quadrillion.

Mugger: Good. Now we will do some maths. Let us say that the 10 livres that you have in your wallet are worth to you the equivalent of one happy day. Let's call this quantity of good 1 Util. So I ask you to give up 1 Util. In return, I could promise to perform the magic tomorrow that will give you an extra 10 quadrillion happy days, i.e. 10 quadrillion Utils. Since you say there is a 1 in 10 quadrillion probability that I will fulfil my promise, this would be a fair deal. The expected Utility for you would be zero. But I feel generous this evening, and I will make you a better deal: *If you hand me your wallet, I will perform magic that will give you an extra 1,000 quadrillion happy days of life.*

Pascal: I admit I see no flaw in your mathematics.

Mugger: This is my final offer. You're not going to pass up a deal that we have just calculated will give you an expected Utility surplus of nearly 100 Utils, are you? That's the best offer you are likely to see this year.

Pascal: Is this legitimate? You know, I've committed myself to trying to be a good Christian.

Mugger: Of course it's legitimate! Think of it as foreign trade. Your currency is worth a lot in the Seventh Dimension. By agreeing to this transaction, you give a major boost to our economy. Oh, and did I mention the children? If only you could see the faces of the sweet little orphans who will be made so much better off if we get this influx of hard currency – and there are so many of them, so very, very, very many . . . .

Pascal: I must confess: I've been having doubts about the mathematics of infinity. Infinite values lead to many strange conclusions and paradoxes. You know the reasoning that has come to be known as 'Pascal's Wager'? Between you and me, some of the critiques I've seen have made me wonder whether I might not be somehow confused about infinities or about the existence of infinite values . . .

Mugger: I assure you, my powers are strictly finite. The offer before you does not involve infinite values in any way. But now I really must be off; I have an assignation in the Seventh Dimension that I'd rather not miss. Your wallet, please!

Pascal hands over his wallet.

Mugger: Pleasure doing business. The magic will be performed tomorrow, as agreed.[1]

Oxford University
Future of Humanity Institute
Faculty of Philosophy & James Martin 21st Century School
Suite 8, Littlegate House, 16/17 St Ebbe's Street
Oxford OX1 1PT, UK
nick.bostrom@philosophy.ox.ac.uk

1 Related scenarios have recently been discussed informally among various people. Eliezer Yudkowsky named the problem 'Pascal's mugging' in a post on the *Overcoming Bias* blog (*http://www.overcomingbias.com/2007/10/pascals-mugging.html*). I am grateful to Toby Ord and Rebecca Roache for comments.