ABSTRACT. The two envelopes problem has generated a significant number of publications (I have benefited from reading many of them, only some of which I cite; see the epilogue for a historical note). Part of my purpose here is to provide a review of previous results (with somewhat simpler demonstrations). In addition, I hope to clear up what I see as some misconceptions concerning the problem. Within a countably additive probability framework, the problem illustrates a breakdown of dominance with respect to infinite partitions in circumstances of infinite expected utility. Within a probability framework that is only finitely additive, there are failures of dominance with respect to infinite partitions in circumstances of bounded utility with finitely many consequences (see the epilogue).

KEY WORDS: Pair of envelopes, countable additivity, principle of dominance, finite additivity, infinite partitions

Variants of the following problem have acquired a certain renown (see, e.g., Nalebuff, 1989). If offered the choice between a sum of money, \( r \) (you know nothing of the magnitude of \( r \), nor how it was determined), and a gamble on a fair coin in which you win \( 2r \) if the coin comes up heads, \( r/2 \) if tails, standard utility theory has it that, provided you are a risk neutral money-seeker, you should opt for the gamble. Imagine you are offered the choice 100 times. If you consistently opt to take \( r \), you will emerge with a total of \( 100r \); if, on the other hand, you consistently take the gamble, you will, on average, emerge with a total of \( 125r = 50(2r) + 50(r/2) \). Hence, in a single shot, the gamble has expected actuarial value (‘eav’) \( 5r/4 \), which is greater than \( r \).

Suppose, however, that the coin has been tossed, and the winnings, \( w (w = 2r \text{ or } w = r/2) \), depending on the outcome of the toss, you know not which), placed in an envelope, \( W \), which is then handed to you. You are now offered the choice of keeping \( W \), or trading

---

* I have benefited greatly from the comments of Ned McClenen and Teddy Seidenfeld.

it for the envelope $R$ containing the sum $r$. You reason as follows. Imagine 100 trials in which $W$ contains $w$. In 50 of those trials, on average, $r = w/2$; and in 50, $r = 2w$. On average, then,

$$r = \frac{50(w/2) + 50(2w)}{100} = 5w/4$$

Thus $\text{eav}(R) > \text{eav}(W)$, so trading is the thing to do.

What has transpired? In the standard reasoning about the gamble, the sum $r$ is fixed, and $w$, the amount of the winnings, varies, across the imagined trials. In the second line of reasoning, $w$ is fixed, and $r$ varies. There are two different perspectives, as it were: the $R$ (fix $r$), and the $W$ (fix $w$). Adopting both together apparently clashes with standard utility theory: $\text{eav}(R)$ would be both greater than, and less than, $\text{eav}(W)$. Nalebuff (1989, pp.171–172) sees a paradox here – he imagines two players each holding one of the envelopes, and he complains: “Trading envelopes cannot make both participants better off”.

We cannot simply appeal to the conventional wisdom that the $R$ perspective is clearly correct because it leads to the outcome we all know to be superior for the risk-neutral money-seeker – take the winnings. We must ask: why is this outcome superior? We want to respond: because, on average, the winnings exceed the original sum, $r$. But this is simply to presuppose that $r$ is fixed across the relevant imagined trials – i.e., to presuppose the $R$ perspective. And this is to beg the question: fix $w$, vary $r$, and taking $R$ appears superior.

So far, I have given no consideration to the prior probability distribution of $r$. And ignoring prior probabilities, or ‘base rates’, is a common source of probabilistic fallacies (see, e.g., Kahneman, Slovic, and Tversky, 1982, Chapter 10). So let us consider the following simple distribution (distribution A):

$$p(r = 2) = p(r = 4) = p(r = 8) = \frac{1}{3};$$

$$p(w = 2^{n+1}|r = 2^{n}) = p(w = 2^{n-1}|r = 2^{n})$$

$$= \frac{1}{2}, \text{ for } n = 1, 2, 3.$$

There are six equally likely pairs, $\langle r, w \rangle$: $\langle 2, 1 \rangle, \langle 2, 4 \rangle, \langle 4, 2 \rangle, \langle 4, 8 \rangle, \langle 8, 4 \rangle, \langle 8, 16 \rangle$. In pairs 2, 4, & 6, $r = w/2$; in pairs 1, 3, & 5, $r = 2w$. So, amongst these pairs, $p(r = (w/2)) = p(r = 2w) = 1/2$. Thus it seems:

$$(\alpha) \text{ For any } w: \text{eav}(R) = (2w)p(r = 2w)$$