

**The Two Envelope Paradox
and Using Variables Within the Expectation Formula**

Eric Schwitzgebel
Department of Philosophy
University of California at Riverside
Riverside, CA 92521-0201
eschwitz at domain- ucr.edu
951 827 4288

Josh Dever
Department of Philosophy
University of Texas at Austin
Austin, TX 78712-1180
512 471 5611
dever at domain- mail.utexas.edu

November 6, 2007

The Two Envelope Paradox and Using Variables Within the Expectation Formula

The present paper presents a diagnosis of what goes wrong in the reasoning in the “closed envelope” version of what is sometimes called “Two Envelope Paradox” or “Two Envelope Problem” or “Exchange Paradox”. Plainly, some constraint on the use of variables within the expectation formula is necessary to escape the paradox. We argue that previous proposed constraints are too restrictive: One can avoid the paradoxical reasoning as long as the conditional expectations of the relevant variables are the same in each event in the partition – provided that all the relevant equations are linear.

Note to editor and referee: We have kept the technical formalities in this paper to a minimum to reach the broadest possible audience, since we’ve found that there’s considerable popular interest in this paradox. If you think a more formal and technical proof would be appropriate for the readership of AJL, we would be happy to provide one.

The Two Envelope Paradox and Using Variables Within the Expectation Formula

The paradox. You are presented with a choice between two envelopes. You know one envelope contains twice as much money as the other, but you don't know which contains more. You arbitrarily choose one envelope – call it Envelope A – but don't open it. Call the amount of money in that envelope X . Since your choice was arbitrary, the other envelope (Envelope B) is 50% likely to be the envelope with more and 50% likely to be the envelope with less. But, strangely, that very fact might make Envelope B seem attractive: Wouldn't switching to Envelope B give you a 50% chance of doubling your money and a 50% chance of halving it? Since double or nothing is a fair bet, double or half is more than fair. Applying the standard expectation formula, you might calculate the expected value of switching to Envelope B as $(.50)\frac{1}{2}X$ [50% chance it has less] + $(.50)2X$ [50% chance it has more] = $(1.25)X$. So, it seems, you ought to switch to Envelope B: Your expected return – your return on average, over the long run, if you did this many times – would seem to be 25% more. But obviously that's absurd: A symmetrical calculation could persuade you to switch back to Envelope A. Hence the paradox.

Where have we gone wrong? What's the flaw in the reasoning? Despite many interesting discussions of alternative ways to reason through the Two Envelope paradox, no one has given a fully adequate answer to this question – no one has fully exposed the nature of the misstep.¹ The problem, surely, has something to do with how variables are deployed in the fallacious argument. Proper diagnosis of the fallacy, then, should help clarify more generally what counts as proper or improper use of variables within the expectation formula.

Other discussions of the Two Envelope paradox have tended either to focus on the “open envelope version” of the paradox, in which one gets to see the contents of the chosen envelope before deciding whether to switch (and we agree with the general consensus here that whether to switch depends on what you see, and only very weird probability distributions generate the result that you should switch no matter what you see); or they have satisfied themselves with vague remarks the mathematical grounding of which is unclear; or they have advocated constraints on the use of variables in the expectation formula that are, we think, considerably more restrictive than necessary.

An analogously absurd case. Our solution to the paradox essentially analogizes the reasoning above to the following reasoning, where the source of the problem is more obvious: You are presented with an envelope containing either \$1, \$2, \$10, or \$20 with equal probability. You are given the choice between two wagers. On the first, you receive twice the amount of money in the envelope, if the amount in the envelope is \$1 or \$2, or just the amount of money in the envelope if the amount in the envelope is \$10 or \$20. The second wager is the reverse: You receive twice the amount of money in the envelope if the envelope contains \$10 or \$20 and just the amount of money in the envelope if it contains \$1 or \$2. Assigning X to the amount in the envelope, you reason that on either bet there is a 50% chance you will receive X and a 50% chance you will receive $2X$ (for an expectation of $3/2 X$), so you are indifferent between the two bets.

Wager 1	Wager 2
$\$1 * 2$	\$1
$\$2 * 2$	\$2
\$10	$\$10*2$

\$20

\$20*2

Clearly, however, the second wager is preferable. It's much better to have the chance to double \$10 or \$20 than to have the chance to double \$1 or \$2. The proposed fallacious calculation is fallacious because it does not take that into account. (The actual expectation, which can be calculated on a case-by-case basis, is \$9 for the first wager and \$15.75 for the second.) In Wager 1, the expected value of X in the "2X" part of the formula is much lower than the expected value of X in the "X" part of the formula; in Wager 2, the reverse is the case. A decision-theoretic calculation in which a random variable does not maintain the same expected value in each of its occurrences has no guarantee of producing proper results.

The Solution. Analogously, in the Two Envelope Paradox, the expected value of X in the "2X" part of the formula (where Envelope A is the envelope with less) is less than the expected value in the " $\frac{1}{2}X$ " part of the formula (where Envelope A is the envelope with more).² You would expect less in Envelope A if you knew that it was the envelope with less than you would if you knew it was the envelope with more. Allowing X to have different expectations in different parts of the formula in this way is like comparing apples and oranges. The "X" in the "2X" just isn't the same as the "X" in the " $\frac{1}{2}X$ " part.

The proper course of action in the Two Envelope Paradox can be non-paradoxically calculated by setting X to the amount in the envelope with less and calculating the expected value of Envelope B as $(.50)X + (.50)2X = \frac{3}{2} X$ – and the expected value of Envelope A likewise as $(.50)X + (.50)\frac{1}{2}X = \frac{3}{4} X$. In these calculations the expectation of X in the first term of each equation is identical to its expectation in the second term: The expected amount of money in the envelope with less does not change depending on

whether Envelope A is the envelope with more or Envelope B is. The availability of such a non-paradoxical calculation is old news, of course; the novelty here is the identification of the crucial difference between the paradoxical and non-paradoxical calculation.

In general, we propose as a constraint on the use of variables within the expectation formula that their expected value be the same at each occurrence in the formula. More technically: For all events A_i in the partition of the outcome space, $E(X/A_i) = E(X)$. Abiding by this constraint guarantees the legitimacy of calculations using X as a variable, if all the equations involved are linear (as we will demonstrate more fully below).³

Stronger constraints are too restrictive. Jackson and Oppy (1994) and others following them have proposed a stronger constraint: that to use a formula like $E(Y) = (.50)\frac{1}{2}X + (.50)2X$ it must be the case that for all values of X there's a 50% chance that the value of Y is half the value of X and a 50% chance that the value of Y is twice the value of X .

While applying this constraint would indeed allow one to avoid the paradox, it also rules out other cases where the formula seems intuitively appropriate. Suppose for example that you're about to mug Mary. Around the corner comes someone else – either Terri or Geri, with 50% likelihood of each. You know that Terri usually carries about half as much money as Mary and Geri usually carries about twice as much. It's perfectly appropriate (moral remonstrances aside) to calculate the expected value of letting Mary go in favor of mugging the oncoming party as $(.50)\frac{1}{2}X + (.50)2X$. To calculate in this way, it is not necessary that for all possible dollar amounts in Mary's purse, there be a 50% likelihood that the person coming around the corner has half as much and a 50% likelihood she has twice as much. Perhaps when Mary has \$84.57 (which can't even be

halved), Terri always has \$101.23. Maybe Geri sometimes has the same amount as Mary, sometimes four times as much, and never exactly twice as much – as long as on average she has twice as much, the calculation works, accurately reflecting the long-run expectations. What matters is not that the relationships among the each particular possible value of X and Y exactly mirror the relationships in the overall formula, but rather that the overall expected values of X and Y exhibit the right relationship.

In principle, of course, the expectations could be calculated case-by-case for different possible values carried by Mary, and some purists we've encountered insist that calculating case by case is the only "technically correct" approach – that one simply cannot legitimately combine random variables in the way suggested. The problem with this purism, of course, is that case-by-case calculation may often in practice be difficult or impossible. Thus, it's of potentially great value to the decision theorist to know when case-by-case calculation is genuinely necessary, and when it may be circumvented by short-cut techniques without affecting the outcome of the decision – which is, of course, exactly the question the Two Envelope paradox raises so forcefully.

Needless to say, we see little value in still stronger constraints, such as (per Jeffrey 1995) that one can discharge such X-for-Y substitutions only when X is a true constant. Such excess caution needlessly robs us of the convenience of simple calculations.

Generalizing. Abiding only by our constraint allows also us to generalize to other cases, less intuitive, that stronger constraints forbid us. Consider this case: You have a choice between two wagers. In the first wager, a fair coin is tossed. If it lands heads, you are to draw one of three cards, marked 0, 2, and 4, winning half the amount on the card. (i.e., \$0, \$1, or \$2). If it lands tails, you are to draw one of two cards, marked 1 and 3,

winning two more than the amount on the card (i.e., \$3 or \$5). The second wager begins with a similar coin flip and drawing. However, given heads you win 70% of the amount on the card, plus 1 (\$1, \$2.40, or \$3.80). Given tails you win simply 70% of the amount on the card (\$0.70 or \$2.10).

Wager A:	Heads: 0 → \$0	Tails: 1 → \$3
	2 → \$2	3 → \$5
	4 → \$4	
Wager B:	Heads: 0 → \$1	Tails: 1 → \$0.70
	2 → \$2.40	3 → \$2.10
	4 → \$3.80	

We can let X be the amount on the card: The expectation of X is the same given heads or tails – 2 in both cases. The first wager is thus worth $(.50)½X + (.50)(X+2)$, which simplifies to $(.75)X + 1$. The second wager is worth $(.50)[(.7)X+1] + (.50)(.7)X$, which simplifies to $(.70)X + .5$. We can thus see that the first wager is preferable without calculating case-by-case – which is obviously a great advantage as the number of cards in the two decks increases! Stronger constraints forbid such calculations.

The proof. As long as one abides by the constraint we propose – that the conditional expectation of the variable be the same in each term or condition of the equation (i.e., in each event in the partition) – and by one additional constraint, that the functions be linear (this second constraint, though necessary, is perhaps not obvious), one will come to the same results in one’s calculations as one would working by the more arduous case-by-case method, calculating the expectations for each particular value. Why? If the expectation of Y (the ultimate outcome you’re interested in) is a linear function $g_i = m_i x + b_i$ of the expectation of X (the variable in question) in various conditions A_i (possibly a different linear function in different A_i), then

$$E(Y) = \sum_i [m_i E(X/A_i) + b_i] P(A_i).$$

If X has the same expected value in the different conditions A_i , then $E(X/A_i) = E(X)$, and consequently

$$E(Y) = \sum_i [m_i E(X) + b_i] P(A_i).$$

In other words, one can calculate the expectation of Y by summing the different g_i functions on the expectation of X (which needn't actually be calculated) times the probability of the A_i – the kind of stuff we were doing above, the kind of maneuver we'd like to make, that it often makes intuitive sense to make, but that the Two Envelope paradox may bring us to doubt the validity of. Getting rid of the $E(X/A_i)$ in favor of $E(X)$ is crucial here: It means one can treat X as the same in every condition, which is key to simplifying the equation into an interpretable result (e.g., simplifying $(.50)X + (.50)2X$ into $3/2 X$). The linearity is crucial to distributing the g_i functions outside the scope of the expectation of X in the first step.⁴

Conclusion. We don't claim to be presenting a novel or profound mathematical result. But we do hope these remarks will prove useful to the reader who feels the pull of puzzlement, as we do, about what has gone wrong in the reasoning of the Two Envelope Paradox but sees no straightforward solution that doesn't – as do all published solutions we've seen – forbid other sorts of calculations that it seems perfectly reasonable to make.⁵

¹ Discussions include Nalebuff 1989; Marinoff 1993; Jackson, Menzies, and Oppy 1994; Broome 1995; Chihara 1995; Jeffrey 1995; Rawling 1997; Clark and Shackel 2000; Horgan 2000; Chalmers 2002; Meacham and Weisberg 2003; Priest and Restall 2003; Dietrich and List 2004; Langtry 2004.

² We assume, contra Langtry (2004), that X here is a random variable with an expected value. If that seems troublesome to you in the case as described, imagine the following variation: For every possible value of X from one cent to \$20 trillion, you have some (small!) determinate degree of confidence that that value is the value in X . For higher and lower values, your subjective probability is zero. Perhaps, indeed, the person offering you the wager, in whom you have absolute faith, provides you with a full list of those probabilities. Suppose, also, that you also have a ballpark sense of how much richer you'll be if you take the contents of the envelope: probably just a few dollars. We see no reason, in such a case – or indeed in the more sparsely presented case – not to suppose, for decision-theoretic purposes, that X can be interpreted as a random variable with a determinate or approximate expected value.

³ Chihara (1995) and Horgan (2000) have proposed constraints that bear some similarity to ours, but which in our view are vague and difficult to interpret, and which fail to note the importance of linearity. Priest and Restall (2003) also offer a similar constraint, which is a bit clearer but too strong: that the actual value of X be the same in both events. That this constraint is needlessly strong can be seen from the example involving the coin flip and cards below, and from the proof.

⁴ Consider the following case, where the violation of linearity is to blame: A coin is flipped. If it lands heads, \$10 is put in an envelope. If it lands tails, either \$0 or \$20 is put in the envelope. You are given a choice of the following two wagers: (1.) The amount of money in the envelope, if the coin landed heads, or the amount squared if it landed tails, or (2.) the amount of money in the envelope, if the coin landed tails, or the amount squared if it landed heads. The expectation of X is the same, given heads or tails, but linearity is violated, and characterizing your expectation as $.5X + .5X^2$ in the two cases leads to an erroneous recommendation of indifference. No such trouble if instead of squaring, one doubles and adds one.

⁵ The outlines of this position were developed in 1993, with input from numerous people in the Berkeley Philosophy Department, most memorably Charles Chihara, Edward Cushman, and Sean Kelly. We have also profited from more recent discussions with Terry Horgan, Brian Skyrms, Peter Vanderschraaf and others.

References:

- Broome, J. (1995). The two envelope paradox. Analysis 55, 6-11.
- Chalmers, D.J. (2002). The St. Petersburg two-envelope paradox. Analysis 62, 155-157.
- Chihara, C.S. (1995). The mystery of Julius: A paradox in decision theory. Philosophical Studies 80, 1-16.
- Clark, M., and Shackel, N. (2000). The two envelope paradox. Mind 109, 415-442.
- Dietrich, F., and List, C. (2004). The two-envelope paradox: An axiomatic approach. Mind 114, 239-248.
- Horgan, T. (2000). The two envelope paradox, nonstandard expected utility, and the intensionality of probability. Noûs 34, 578-603.
- Jackson, F., Menzies, P., and Oppy, G. (1994). The two envelope paradox. Analysis 54, 43-45.
- Jeffrey, R.C. (1995). Probabilistic thinking. Available at <http://www.princeton.edu/~bayesway/ProbThink/TableOfContents.html> .
- Langtry, B. (2004). The classical and maximin versions of the two envelope paradox. Australasian Journal of Logic 2, 30-42.
- Marinoff, L. (1993). Three pseudo-paradoxes in 'quantum' decision theory: Apparent effects of observation on probability and utility. Theory and decision, 35, 55-73.
- Meacham, C.J.G., and Weisberg, J. (2003). Clark and Shackel on the two-envelope paradox. Mind 112, 685-689.
- Nalebuff, B. (1989). Puzzle: The other person's envelope is always greener. Journal of Economic Perspectives 3, 171-181.

Priest, G., and Restall, G. (2003). Envelopes and indifference. Available at
<http://consequently.org/papers/envelopes.pdf> .

Rawling, P. (1997). Perspectives on a pair of envelopes. Theory and Decision 43, 253-
277.