What’s in the two envelope paradox?

Alexander D. Scott & Michael Scott

You are given a choice between two envelopes. Both envelopes contain cheques, one twice the value of the other, but you do not know the value of either cheque or which cheque is in which envelope. Having chosen one envelope, you are then given the opportunity to change to the other envelope. Given that your objective is to maximize your wealth, should you refuse the offer or swap envelopes? Suppose the amount of the cheque in the chosen envelope is $x$. Then the other envelope contains either $x/2$ or $2x$, and these two possibilities are equally likely. It follows that the expected value of the cheque in the other envelope is

\[
\frac{1}{2} \cdot \frac{x}{2} + \frac{1}{2} \cdot 2x = 1.25x
\]

So it seems that you should swap because the other envelope will on average contain a larger sum. But it also seems that this conclusion must be false. For there can be no way of deducing some advantage in choosing one envelope over another without some information about the contents of the envelopes, in addition to the fact that the cheque in one envelope is twice the value of the cheque in the other. Choosing one envelope cannot put you in the position of reasonably expecting that the other contains more, for had you chosen the other envelope first the same argument would lead you to the opposite conclusion.¹ In this paper we will indicate certain shortcomings in earlier assessments of the stated paradox and present a revised version of the paradox for which swapping envelopes is the best option. We will be concerned throughout with the possibility of a probabilistic argument in favour of swapping envelopes.

The argument for swapping envelopes requires that, whatever amount $x$ is in the envelope you pick first, you are equally likely to find $x/2$ or $2x$ in the other envelope. Jackson, Menzies and Oppy (1994) observe that no such assumption is given in the statement of the paradox, and argue moreover that the argument ‘is false, at least for any rational, minimally informed person’. For example, given the reasonable assumptions that there is a finite amount of money available and that there is a smallest unit of currency, there are values of $x$ for which it is unlikely that the other envelope contains $x/2$ and values for which it is unlikely that it contains $2x$. But there is a more immediate way of disposing of the paradox. Let $X$ be the amount in the envelope you first pick and let $L$ be the larger of the

¹ The argument would not, however, lead you to change back to the first envelope after making the swap, as John Broome (1995) suggests, unless you ignore the earlier conclusion that the expectation for the value of the cheque in the ‘other’ envelope (which you now hold) is $1.25x$.
amounts contained in the two envelopes. The argument for swapping envelopes requires that, for every $x$,

\[ P(X=x \text{ and } L=x) = P(X=x \text{ and } L=2x). \]

We shall show that this assumption entails a contradiction.

Since, for every $x$,

\[ P(X=x \text{ and } L=x) = 0.5 \times P(L=x) \]

and

\[ P(X=x \text{ and } L=2x) = 0.5 \times P(L=2x) \]

we must have

\[ P(L=x) = P(L=2x). \]

Now suppose that $P(L=x) > 0$ for some $x$. Then

\[ P(L=2x) = P(L=x) > 0 \]

and

\[ P(L=4x) = P(L=2x) > 0 \]

and so on. That is, the events $L=x$, $L=2x$, $L=4x$, all have the same positive probability and so the sum of their probabilities is infinite, which is a contradiction. Therefore we must have $P(L=x) = 0$ for every $x$. But this proves equally problematic. If we take the distribution of $L$ to be given by a probability density function $f(x)$, our first thought might be that the argument requires $f(x) = f(2x)$ for every $x$. However, this proves to be a mistake. Let us write $S$ for the smaller of the amounts in the two envelopes. Then the best interpretation of the argument seems to be that $S$ and $L$ should have the same distribution (this can be expressed more formally in terms of conditional probability). This implies that $f(2x) = 0.5f(x)$ for (almost) every $x$. But, as Castell and Batens (1994) prove, this is not satisfied by any probability density function (and, incidentally, we could not demand $f(2x) = f(x)$ either).

Thus the simple form of the paradox can be rejected without considering pragmatic restrictions on the availability of currency; indeed, without specifying any particular assumption you make regarding distributions, since the argument itself makes an assumption which entails a contradiction.

The paradox has recently been revived by Broome (1995), who shows that there are distributions $L$ such that whatever amount $x$ is in the first envelope, the expected sum in the other envelope is larger. As an example,

---

2 The distribution of $L$ will in general be given by a probability measure $\mu$. However, for the sake of simplicity we restrict our attention to distributions with a density function; our arguments extend straightforwardly to the general case.
Broome gives the distribution where $L = 2^n$ with probability $2^{(n-1)/3^n}$, for $n = 1, 2, 3, \ldots$, and $L$ takes any other value with probability zero. Broome correctly calculates that, given that the envelope you pick contains $x$, the expected sum contained in the other envelope is $11x/10$. Thus for his distribution the conditional expectation for $Y$ given that the first envelope contains $x$ satisfies $E(Y|X=x) > x$ for every $x$ with $P(X=x) > 0$, where we write $Y$ for the contents of the other envelope. We call such a distribution paradoxical.

The existence of paradoxical distributions seems at first to lead us back to square one: for any sum $x$ in the envelope you pick, you expect to be better off by changing. But suppose $X$ is a paradoxical distribution on the positive integers. Then $E(Y|X=x) > x$ for every $x$ with $P(X=x) > 0$, which can easily be shown to be equivalent to the condition that $P(X=2x) > 0.5 \times P(X=x)$ for every $x$. Castell and Batens show that this implies that $X$ has infinite expectation and then argue that the appearance of paradox ‘derives from the fact that if a random variable has infinite expectation, then its true value is bound to be less than its expectation value’. They proceed to dismiss infinite expectation as an absurdity (a surprising conclusion, since they object to the claim of Jackson, Menzies and Oppy that distributions with unbounded domains are ‘fantastical’ because ‘it is precisely in these “fantastical” cases that the Two Envelope Paradox is a serious conceptual challenge’). In contrast, Broome claims that ‘strictly, a distribution with no finite mean does not have a mean that exceeds any finite amount; it has no mean at all’. So although Broome does not find an infinite expectation for $X$ absurd, he claims that properly speaking we should say that $X$ does not have an expectation. Undeterred by this restriction, Broome advances an argument for switching which does not, he asserts, refer to the mean of the distribution: given a paradoxical distribution of values, the first envelope contains a ‘fixed but unknown’ amount $x$, and whatever its value, we expect to be better off taking the second envelope. Hence we should swap envelopes. However, this is mistaken: to conclude that $E(Y-X) > 0$ (i.e. we want to swap) from the premiss that $E(Y|X=x) > x$ for every $x$ with $P(X=x) > 0$ (i.e. for every $x$, we want to swap) we must take an average. Broome conceals a reference to an average case by calling $x$ a ‘fixed but unknown’ quantity. Of course, $x$ does have a fixed value in the sense that there is a particular amount in each envelope, but to argue that, irrespective of the particular value of $x$, it is better to swap, we must calculate the mean.

Both Broome and Castell and Batens see problems in distributions without finite expectations, that is, distributions for which the expectation is a

---

3 Note that we are moving from $E(Y-X|X=x) > 0$ to $E(Y-X) > 0$. 

divergent sum (or, in general, a divergent integral; the same argument applies). But they fail to make a crucial distinction between two types of divergence. A series such as $1 + 1 + 1 + 1 + \ldots$ can quite consistently be said to have infinite sum; however, a series such as $1 + (-1) + 1 + (-1) + \ldots$ cannot consistently be said to have any sum at all (see Hardy 1938, ch. IV for a lucid discussion). There is a corresponding distinction between random variables with infinite expectation and random variables with no expectation at all.\(^4\) Castell and Batens consider only the former case, which they dismiss as absurd, while Broome addresses only the latter on the basis that infinite expectation is ill-conceived. But both types of random variable are well-defined and both play a role in the paradox. We can say consistently that $X$ has infinite expectation, but that the amount $Y - X$ gained by switching envelopes has no expectation, since the positive and negative parts of the sum both diverge (informally, we expect both to gain and to lose an infinite amount). Thus Broome’s covert reference is to a random variable with no expectation.

We have now reached a mathematical solution to Broome’s version of the paradox. The expectation for the contents of the envelope you pick is infinite, as is the expectation for the contents of the other envelope. That is, the random variables $X$ and $Y$ both have infinite expectation. However, the amount you gain from switching envelopes does not have an expectation, and thus it makes no sense to have an a priori preference as to which envelope you pick. Even when you have picked an envelope there is no good reason to switch, because until you have actually seen the contents of the envelope you know only that the amount $x$ in the envelope is distributed according to the random variable $X$. Only when you open the envelope and discover the particular value of $x$ that it contains (a necessary condition that is neglected by Broome) can you consistently reason that on average, given this particular sum in the envelope, you are better off switching.

It is now a simple matter to change the conditions of the choice between envelopes to give us a ‘working’ version of the paradox. As before, you are given a choice between two envelopes. Both the envelopes contain cheques chosen according to a given paradoxical distribution, one cheque twice the value of the other, but you do not know the amounts of the cheques or which cheque is in which envelope. Having chosen one envelope, you open it and see the value of the enclosed cheque. You are then given the opportunity to change to the other envelope. Given this setup, you will definitely be better off on average switching envelopes.

\(^4\) Most modern textbooks follow this course, although Feller (1968) disallows infinite expectations. In this case our arguments still apply, except that we cannot talk about a prior expectation.
Although analogous to the initial statement of the two envelopes paradox, our revised version is not puzzling for the same reasons. It seemed before that there could be no good reason for exchanging envelopes, because there was no way of knowing that the envelope that you chose first contained the larger cheque. Prima facie, your chances of choosing the larger or smaller cheque should be equal, and altering the setup by allowing you to see the contents of the first envelope and distributing the values of the cheques in a particular way should not alter this fact. But in the revised version of the paradox, there is a distinction between the probability that you will initially choose the larger or smaller cheque, and the probability – given the paradoxical distribution and the value of the cheque in the first envelope – that the other envelope contains twice as much. The former probability, which can be calculated before the choice is made, indicates that you are on average as likely to choose the larger as the smaller cheque, while the latter probability, which can only be calculated after the first choice has been made and the value of the cheque is known, indicates that for the particular value of the cheque in the first envelope there is a good chance that the other envelope will contain the larger sum. But this does not imply a mysterious facility to choose the envelope with the smaller cheque, for the expected advantage in swapping envelopes can only be given for particular values of the cheque in the first chosen envelope. Before choosing an envelope, no expected gain or loss from swapping can be calculated, because it is impossible to take a mean value for the amount gained by switching. You can predict for any particular value of the cheque in the envelope you choose first that it is reasonably likely that the other envelope contains twice as much, because the possible monetary values are distributed in a specified way, i.e. the different possible pairs of values are weighted. But this is consistent with your expecting no advantage in choosing one or other envelope from the outset. A more puzzling consequence of the revised version of the paradox is that whichever envelope you choose, then given the particular value of the enclosed cheque it will be in your best interests to swap. Thus if the other

---

5 You need not always think you are more likely to have picked the smaller cheque first, but only that, once you have opened an envelope, the probability that you have picked the smaller cheque is enough that you expect to gain from swapping envelopes. In the distribution of Broome given above, unless you find 1 in the envelope you open, you are slightly more likely to have picked the larger amount (given the amount you find).

6 The same argument applies in the case when the amounts in the two envelopes are chosen independently according to a distribution with infinite mean. In this case, neither envelope is preferable before opening but, after opening one envelope, the other envelope is always preferable.
envelope were chosen by another person, you would both find it in your best interests to exchange. This is clearly shown by an analogous version of the two envelope paradox described by J.E. Littlewood (1986: 26), which he attributes to Schrödinger. In Littlewood’s account, we have an indefinite supply of cards: one marked 0 on one side and 1 on the other, ten marked 1 on one side and 2 on the other, and so on (in general, 10^n marked n on one side and n+1 on the other). A card is drawn at random and held between two players A and B so that each sees just one side. The players now bet with each other (or with a bookie, at evens) and the player with the lower number wins. Now suppose that A sees 1. Since there are ten times as many cards marked n and n+1 as there are marked n-1 and n, A has probability 10/11 of winning; B also has probability 10/11 of winning, by the same argument.

This version of the paradox fails because of the impossibility of (to use Littlewood’s phrase) the ‘monstrous hypothesis’ that each card is equally likely. However, we can resuscitate the paradox as follows. A card is chosen with \(x\) on one side and 2x on the other, where \(x\) has a paradoxical distribution. The card is shown to A and B so that each sees one side, and A and B are given the opportunity to swap. Then both A and B can correctly reason that it would be preferable to receive the sum on the other side of the card.

This apparently paradoxical conclusion can now be seen to be a consequence of a fact about series without a sum. Given a series such as

\[
1 + (-2) + 3 + (-4) + \ldots
\]

we can bracket the sum as

\[
(1-2) + (3-4) + \ldots
\]

or as

\[
1 + (-2+3) + (-4+5) + \ldots
\]

In the former case, the series would appear to sum to \(-\infty\), while in the latter it would appear to sum to \(+\infty\); in fact, since both positive and negative parts diverge, the series does not have a sum. Now writing A’s side of the card first, with Broome’s distribution we get 1,2 or 2,1 with probability 1/6 each, 2,4 or 4,2 with probability 1/9 each, and so on (2^n,2^{n+1} or 2^{n+1},2^n with probability 2^{n-1}/3^n each). Thus the expected gain if A switches is

\[
(1/6)(1) + (1/6)(-1) + (1/9)(2) + (1/9)(-2) + \ldots
\]

The argument that A should choose to swap without looking at the card is equivalent to grouping together 2,1 and 2,4, then 4,2 and 4,8, and so on, according to the amount observed by A. Thus the expectation is calculated as
The equivalent argument for B calculates the same sum (the loss for B) as

\[
\left(\frac{1}{6}\right)(-1) + \left(\frac{1}{6}\right)(1) + \left(\frac{1}{9}\right)(-2) + \left(\frac{1}{9}\right)(2) + \frac{2}{27}(2) + \ldots
\]

In the former sum, all terms are positive and the argument concludes that A gains on average by swapping; in the latter sum, all terms are negative, and the opposite conclusion is reached. In fact, the sum does not have a value, and neither argument holds. However, in the case that A does look at the card, the situation is different. A simple calculation shows that, after looking at the card, A should choose to swap.

Although our argument in this paper has focused only on integer-valued distributions, it can be made for other probability distributions. In order to show that a given distribution is paradoxical, we must calculate a certain conditional probability. However, our intuitions about finite distributions (such as coins, dice and cards) can mislead when extended to arbitrary probability distributions.\(^7\) Since \(P(X=x) = 0\) for every \(x\), the distribution of \(Y\) conditional on \(x\) is not defined. It therefore becomes necessary to define a new notion of conditional probability. It turns out that there need not be a uniquely defined conditional probability distribution, and that two versions of the conditional probability need only agree almost everywhere (see Billingsley 1986). Castell and Batens attempt to show that no distribution with a density function can be paradoxical. However, the expression that they quote for (conditional) expectation is incorrect; they conclude that for a distribution to be paradoxical it is necessary that \(f(x) \geq 0.5f(x/2)\) for every \(x\), which they show is impossible. The correct condition is \(f(x) \geq 0.25f(x/2)\), which is easily satisfied (Broome gives an example). They rightly observe that the expected profit from switching envelopes can be divergent. But the integral they write down can be infinite, whereas properly stated it has no value. Broome attempts to avoid these problems by refusing to admit infinite expectations and by deriving the conditional expectation of \(Y\), given that \(X=x\), as a limit of events with positive probability (see Appendix A of his paper). However, his derivation of conditional expectation is inadequate: he does not justify his method and incorrectly asserts that the approximation he gives for \(P(y=2x \mid z \leq x \leq z+\Delta z)\) must become precise as \(z\) tends to zero (we could for instance have \(P(z \leq x \leq z+\Delta z) = P(z/2 \leq x \leq (z+\Delta z)/2) = 0\), but \(f(z) > 0\) and \(f(z/2) > 0\)). Broome’s rejection of infinite expectations is legitimate, although as we have argued it is unnecessary. Indeed, the arguments we have made for integer-valued distributions carry over without difficulty to general probability distributions.

\(^7\) In this sense, Jackson, Menzies and Oppy are correct to point out that only certain types of distribution occur in practice; we could add to these those distributions met in a course on elementary probability, which serve to form or solidify our intuitions.
tions (with appropriate definitions of conditional probability and expectation). Nevertheless, we have concentrated on the case when the sums in the envelopes are integers, since this avoids mathematical technicalities and the paradox appears to arise most clearly in this case.

Trinity College, Cambridge
ads@pmms.cam.ac.uk

Department of Philosophy of Religion and Ethics
University of Utrecht
michael@plumpton.demon.co.uk

References