The Two-envelope Paradox

MICHAEL CLARK AND NICHOLAS SHACKEL

Previous claims to have resolved the two-envelope paradox have been premature. The paradoxical argument has been exposed as manifestly fallacious if there is an upper limit to the amount of money that may be put in an envelope; but the paradoxical cases which can be described if this limitation is removed do not involve mathematical error, nor can they be explained away in terms of the strangeness of infinity. Only by taking account of the partial sums of the infinite series of expected gains can the paradox be resolved.

1. Finite distributions

You are presented with two sealed envelopes, one of which contains twice as much money as the other, and you select one at random. You are then offered the chance to swap and take the other instead. If your selected envelope contains x, and your swap is lucky, you get $2x$, but if you are unlucky you get $^1/_2x$. So it seems that your expected utility if you swap is $^1/_2 \times 2x + {}^1/_2 \times {}^1/_2x$, which is $1{}^1/_4x$. For example, if you have £2 in your chosen envelope the other envelope must have either £1 or £4, average £2.50. So it looks as if you should swap. However, exactly the same argument would have been available if you had picked the other envelope in the first place.

As it stands it is not difficult to see what is wrong with this argument. But it can be developed into a paradox that is not so easily resolved.

We shall focus on cases with integral sums of money in the envelopes, which is how money comes in our world.[1] There will be no loss of generality if we take the values to be the powers of 2: 2^0, 2^1, ... , 2^n, 2^{n+1}, ... Let the smallest monetary unit (say £1) be 1: every other unit has to be divisible by 2 without remainder. In its original version the paradoxical argument is plainly fallacious. It is given that you choose randomly between the two envelopes, so that the chance of picking each is 0.5. But it does not follow from the fact that (x, $2x$), ($2x$, x) are equiprobable that (x,$^1/_2x$), (x, $2x$) are equiprobable (cf. Jackson, Menzies and Oppy 1994), as is plain from Fig. 1 below.

[1] As far as we can see, no extra issues of significance are raised by continuous distributions. See the final section of the Appendix below.

Clearly, if you have 1 in your envelope you can only gain by swapping. If you have 2, you get either 1 or 4 on swapping. There is nothing in the description of the set-up that implies that 4 is as likely as 1. But what if the respective probabilities, p_0 and p_1 are such that the *weighted* average of 1 and 4 exceeds 2, and moreover this is true for every value in your envelope greater than 2 (or even on average)?

selected envelope		other envelope	
2	$\Rightarrow$	$1 \times p_0$	Expectation = $\frac{p_0 + 4p_1}{p_0 + p_1}$.
	$\Rightarrow$	$4 \times p_1$	

If $p_1 > \frac{1}{2} p_0$ then $\frac{p_0 + 4p_1}{p_0 + p_1} > \frac{3p_0}{1.5p_0}$. So expectation > 2.

However, it is easy to see from Fig. 1 that, if there is a maximum sum that can be put in an envelope, the average expected gain from swapping is zero, so that not only will no probability distribution for finitely many powers of 2 satisfy the condition that $\forall n p_{n+1} > p_n$ but it will not even yield a gain on average. Each possible gain is matched by an equivalent equiprobable loss. It is also clear that, if you have the maximum sum in your envelope, then you are bound to lose on swapping, which alone is sufficient to show that swapping will not always produce a gain.

In Fig. 1[2] the maximum value in your envelope is 16, with a maximum total sum of 24, 16 in your envelope and 8 in the other; but obviously the situation is the same whatever the maximum value.

A is a random variable for the amount in the selected envelope, and *B* a random variable for the amount in the other envelope. "$+2^0 p_0$" is the gain, weighted by its probability, from swapping an envelope containing 1 for the other envelope when it contains 2; "$-2^0 p_0$" is the loss of 1, weighted by its probability, from swapping an envelope containing 2 for the other envelope when it contains 1, and so on.

Let the probability that 2^n is the smaller sum be $2p_n$. So the probability that the outcome is (1, 2) is p_0 and the probability that the outcome is (2, 1) is also p_0, and, in general, the probability of $(2^n, 2^{n+1})$ = the probability of $(2^{n+1}, 2^n) = p_n$.

Although it is obvious enough from Fig. 1 that the average net gain on swapping is zero, it is worth examining the matter more closely in order to prepare the ground for the genuinely problematic cases.

[2] After Kraitchik (1943, pp. 133–34).

Fig. 1

A \ B	2^0	2^1	2^2	2^3	2^4
2^0		$+2^0p_0$			
2^1	-2^0p_0		$+2^1p_1$		
2^2		-2^1p_1		$+2^2p_2$	
2^3			-2^2p_2		$+2^3p_3$
2^4				-2^3p_3	

When A is 1, the expected gain is $2 - 1$, $= 2^0$. When A has its highest value, 2^{k+1}, the expected gain is $2^k - 2^{k+1}$, $= -2^k$; in other words there is an expected loss of 2^k. Otherwise, $E(B|A) - A$ is the weighted average value of B when A is 2^n, minus 2^n, that is,

$$\frac{2^{n-1}p_{n-1} + 2^{n+1}p_n}{p_{n-1} + p_n} - 2^n.$$

To calculate the average or expectation of $E(B|A) - A$, namely $E(E(B|A) - A)$, we need to weight each instance by its probability, p_0 for $A = 1$, p_k for $A = 2^{k+1}$, and $p_{n-1} + p_n$ for all other values of A.

For example, in the case where $A = 2^1$, $E(B|A) - A = (2^0p_0 + 2^2p_1)/(p_0 + p_1) - 2^1$. Multiplying by the probability weighting, $p_0 + p_1$, we get $2^0p_0 + 2^2p_1 - 2^1p_0 - 2^1p_1$, which equals $2^1p_1 - 2^0p_0$.

In general, $2^{n-1}p_{n-1} + 2^{n+1}p_n - 2^np_{n-1} - 2^np_n = 2^np_n - 2^{n-1}p_{n-1}$.

The sum for the (positive and negative) gains on swapping is tabulated below for our small illustrative example:

2^0p_0	
$+ (2^1p_1$	$- 2^0p_0)$
$+ (2^2p_2$	$- 2^1p_1)$
$+ (2^3p_3$	$- 2^2p_2)$
	$- 2^3p_3$

Since this is a finite sum of finite values its terms may be reordered at will. Move the second column up, and it is evident that each term in the reordered series sums to zero.

$(2^0p_0$	$- 2^0p_0)$
$+ (2^1p_1$	$- 2^1p_1)$
$+ (2^2p_2$	$- 2^2p_2)$
$+ (2^3p_3$	$- 2^3p_3)$

Generalising from this small example: for finite distributions, where the largest total sum in the two envelopes is 3.2^k, the average or expected expected-gain, $E(E(B|A) - A)$, is

$$2^0p_0 + \sum_{n=1}^{k} (2^np_n - 2^{n-1}p_{n-1}) - 2^kp_k$$

$$= 2^0p_0 + \sum_{n=1}^{k} (-2^{n-1}p_{n-1} + 2^np_n) - 2^kp_k$$

$$= 2^0p_0 - 2^0p_0 + 2^1p_1 - 2^1p_1 + \ldots + 2^{k-1}p_{k-1} - 2^{k-1}p_{k-1} + 2^kp_k - 2^kp_k$$

= 0, since all the terms cancel out.

An alternative form of the paradox can be given by comparing the expected gain, given *A*, on swapping with the expected gain, given *B*, on sticking: you go on to argue that the other envelope contains *x*, so that yours has either $^1/_2x$ or $2x$. It then appears profitable both to swap and to stick. But just as the expected gain on swapping in the case of a finite distribution is zero, so is the expected gain on sticking. For $E(B - E(A|B))$, the expected gain given *B*, just interchange + and – in the calculations above:

$$-2^0p_0 + 2^0p_0 - 2^1p_1 + 2^1p_1 - \ldots - 2^{k-1}p_{k-1} + 2^{k-1}p_{k-1} - 2^kp_k + 2^kp_k = 0.$$

In short, if there is an upper bound to the amount in the envelopes, the probabilities cannot give rise to the paradoxical argument.

2. Finite mean expectation for an envelope with an infinite probability distribution

But what if there is no upper limit to the value in an envelope, so that the probability distribution is over a countably infinite set of values? If the average expectation for an envelope is finite, then the result above can easily be extended. Again we can prove that the average expected gain must be zero.

Since we shall now be dealing with sums of infinite series, it may be helpful to recall Zeno's paradox of the Racecourse. In order to run a certain distance Achilles must first run half that distance, then half of what remains, then half of that, and so on ad infinitum. The infinite sequence of these intervals, $^1/_2$, $^1/_4$, $^1/_8, \ldots$, $(^1/_2)^n$, …, ($n \geq 1$), steadily decreases towards 0 as a limit. A sequence which converges to 0 is called a *null sequence*. As a matter of definition, the sum of an infinite series is the limit, *L*, of the sequence of its partial sums, if that limit exists. In other words, the sum is that number *L* such that, for any interval ε, however small, after a certain point in the sequence of partial sums all subsequent members will be within ε of *L*. The infinite series $^1/_2 + ^1/_4 + ^1/_8 + \ldots + (^1/_2)^n + \ldots$ sums to 1, since the sequence of its partial sums, $\langle s_n \rangle = ^1/_2, ^3/_4, ^7/_8, \ldots$, converges towards 1 as *n* tends to infinity. In general, an infinite series has a sum iff the sequence of its partial sums converges. A necessary condition for the sequence of partial sums to converge is that the sequence of the terms of the series is null, which following Haggarty (1993, p. 104) we may call *the vanishing condition.*

To return to the envelopes: the average expectation for your envelope, $E(A)$, is $2^0p_0 + \sum_{n=1}^{\infty}(2^np_{n-1} + 2^np_n)$. Under the supposition that it has a finite value, this must therefore be a convergent infinite series.

The k^{th} partial sum of $E(E(B|A) - A)$, the average expected gain given A, is

$$
\begin{aligned}
&2^0p_0 + \sum_{n=1}^{k}(2^np_n - 2^{n-1}p_{n-1}) \\
&= 2^0p_0 + \sum_{n=1}^{k}(-2^{n-1}p_{n-1} + 2^np_n) \\
&= 2^0p_0 - 2^0p_0 + 2^1p_1 - 2^1p_1 + \ldots \\
&\qquad - 2^{k-2}p_{k-2} + 2^{k-1}p_{k-1} - 2^{k-1}p_{k-1} + 2^kp_k \\
&= 2^kp_k \text{ (since the previous terms cancel out).}
\end{aligned}
$$

Since the sequence of partial sums of $E(E(B|A) - A)$ is the sequence $\langle 2^kp_k \rangle$, we have

$$E(E(B|A) - A) = \lim_{k \to \infty} 2^kp_k.$$

It is worth remembering this equality, since whenever we are discussing the behaviour of the sequence $\langle 2^kp_k \rangle$ or the limit of that sequence we are in effect discussing $E(E(B|A) - A)$, that is to say, the average expected gain on swapping given A.

Now, given that $E(A) = 2^0p_0 + \sum_{n=1}^{\infty}(2^np_{n-1} + 2^np_n)$ is convergent, so is $2^0p_0 + \sum_{n=1}^{\infty} 2^np_n$;[3] therefore by the vanishing condition the sequence of terms $\langle 2^kp_k \rangle$, $k \geq 0$, is null, one which tends to 0 as $k \to \infty$. So the average expected gain, $E(E(B|A) - A)$, is 0, since it is the limit of this very sequence.

[3] See Haggarty (1993, p. 108), first comparison test.

In short, whereas with a finite probability distribution the sequence of partial sums drops to 0 when the complete sum is reached, in the present case the infinite sequence of partial sums converges to 0 as a limit.

Example

As an illustration, here is an example of a countably infinite probability distribution for which the average expected sum in your envelope is finite.

The probability that 1 is the smaller value, $2p_0$, = $^1/_2$. For $n > 0$, the probability, $2p_n$, that 2^n is the smaller value = $(^1/_3)^n$: $^1/_3$, $^1/_9$, $^1/_{27}$, $^1/_{81}$, ... Since $^1/_2 + \sum_{n=1}^{\infty}(^1/_3)^n = {}^1/_2 + {}^1/_2 = 1$, this is a normalised distribution.

The following process will generate this distribution. Select a ball randomly from a pair one of which is white, the other black. If the white ball is drawn let n equal 0. If a black ball is selected on the first draw then select randomly from two white balls and one black one, replacing the black ball if it is drawn, and keep doing this until a white ball is selected. If i is the number of draws preceding the drawing of a white ball, clearly i may take any of the countably many values 0, 1, ..., n, n+1, ... Let n equal i: then set the lower sum to 2^n. The probability that 2^n is the smaller value is just the probability that $n = i$, which in turn is the probability of i black balls and a final white ball, which is $^1/_2$ for $n = 0$, and, for $n > 0$, $^1/_2 \times (^1/_3)^{i-1} \times {}^2/_3 = (^1/_3)^i$. The envelope to contain the smaller sum can then be determined by tossing a fair coin.

The average expectation for an envelope, $E(A)$,

$$2^0 p_0 + \sum_{n=1}^{\infty} (2^n p_{n-1} + 2^n p_n)$$

is $3^3/_4$, which is the limit of the partial sums:

$^1/_4$, $1^1/_{12}$, $1^{105}/_{108}$, $2^{61}/_{108}$, $2^{311}/_{324}$, $3^{217}/_{972}$, ...

For example, the 30th partial sum to 5 decimal places is 3.74997.[4]

[4]
$$2^0 p_0 + \sum_{n=1}^{\infty} (2^n p_{n-1} + 2^n p_n) = p_0 + \sum_{n=1}^{\infty} 2^n p_{n-1} + \sum_{n=1}^{\infty} 2^n p_n$$
$$= p_0 + 2^1 p_0 + \sum_{n=1}^{\infty} 2^{n+1} p_n + \sum_{n=1}^{\infty} 2^n p_n = \tfrac{3}{4} + \sum_{n=1}^{\infty} 3.2^n p_n$$
$$= \tfrac{3}{4} + \sum_{n=1}^{\infty} 3.2^n.\tfrac{1}{2}(\tfrac{1}{3})^n = \tfrac{3}{4} + \sum_{n=1}^{\infty} \tfrac{3}{2}(\tfrac{2}{3})^n = \tfrac{3}{4} + \tfrac{3}{2}.2 = 3\tfrac{3}{4}.$$

The value of the average expected gain given *A*, $E(E(B|A) - A)$, is 0 because the sequence of *its* partial sums, $\langle 2^k p_k \rangle$, is the null sequence: $^1/_4, ^1/_3, ^2/_9, ^4/_{27}, \ldots, {}^1/_2(^2/_3)^k, \ldots$.

What happens here is that the positive expected gains (weighted by their probability), $^1/_4$ when $A = 1$, and $^1/_{12}$ when $A = 2$, are balanced by the infinitely many (weighted) decreasing expected losses when $A > 2$: $-^1/_9, -^2/_{27}, \ldots, -(^1/_9)(^2/_3)^{k-1}, \ldots$, which sum to $-^1/_3$.

Once again, for $E(B - E(A|B))$, the expected expected-gain given *B*, just interchange + and – in the calculations above: as before, the sequence of partial sums $\langle -2^k p_k \rangle$ converges to 0.

So far the results presented, albeit in our own way, have already been established in the literature. Now we move on to the genuinely paradoxical cases.

3. Infinite paradoxical cases

The general proof that $E(E(B|A) - A) = 0$ does not go through when the average expectation for an envelope, $E(A)$, is infinite.

If $E(A)$ is not finite then its series

$$2^0 p_0 + \sum_{n=1}^{\infty} (2^n p_{n-1} + 2^n p_n)$$

must diverge, growing without limit. We cannot therefore infer, as before, that the infinite series $2^0 p_0 + \sum_{n=1}^{\infty} (2^n p_n - 2^{n-1} p_{n-1})$ converges, but neither can we exclude it.[5] In other words, the possibility that $E(E(B|A) - A)$ equals zero is not ruled out. What fails is the general argument that it *must* sum to 0.

When $E(A)$ is infinite, it is possible for the sequence $\langle 2^k p_k \rangle$ to converge to a limit or to be divergent whether by oscillation or by diverging to infinity. To determine which it is we have to look at the particular probability distribution concerned.

[5] This is because the vanishing condition is necessary but not sufficient. The vanishing condition entails that if $E(A)$ converges the sequence of partial sums of $E(E(B|A) - A)$, $\langle 2^k p_k \rangle$, is null, but not that $\langle 2^k p_k \rangle$ is non-null if $E(A)$ does not converge.

We know, from what has already been established in the literature, that there are cases where $\langle 2^k p_k \rangle$ diverges to infinity, and so $E(E(B|A) - A)$ is infinite. Some writers express scepticism about infinite expectations, dismissing them as absurd (Castell and Batens 1994, and cf. Broome 1995):[6] but what if there are cases where the expectation $E(E(B|A) - A)$ has a positive finite value, which will occur whenever $\lim_{k \to \infty} 2^k p_k$ is a positive real number? Such cases would be paradoxical cases in which the average expected gain is uncontroversially well-defined and which an appeal to doing the wrong sort of mathematics or going beyond the sense which our mathematical model provides (abusing infinities) will not dissolve the paradox.[7]

To prove that such cases exist, consider the distribution[8] where the probability that 1 is the smaller value, $2p_0 = {}^1/_6$, and, for $n \geq 1$, the probability, $2p_n$, that 2^n is the smaller value, is $p_{n-1} + ({}^1/_4)^n$.[9]

The expected expected-gain given A will once again be the limit of the sequence of partial sums, $\langle 2^k p_k \rangle$:

$$2^0 p_0 = p_0 = \frac{1}{12}.$$

$$2^1 p_1 = 2^0 p_0 + 2^0 \left(\frac{1}{4}\right)^1 = 2^0 p_0 + \frac{1}{2}\left(\frac{1}{2}\right)^1.$$

[6] Of course they would not say that $E(A)$ is well-defined in these cases either, so they would not say as we have that the expectation for the sum in envelope was infinite, but rather that in these cases there is no finite expectation for A.

[7] In the Appendix below we call such paradoxical cases "best paradoxical", by contrast with cases where the expected gain is not finite, which we call "unbounded paradoxical".

[8] We owe this example to Robert Black.

[9] The distribution is normalised, since

$$\sum_{n=0}^{\infty} 2p_n = 2p_0 + \sum_{n=1}^{\infty} p_{n-1} + \sum_{n=1}^{\infty} \left(\frac{1}{4}\right)^n, \text{ whence}$$

$$\sum_{n=0}^{\infty} 2p_n - \sum_{n=0}^{\infty} p_n = \frac{1}{6} + \frac{1}{3}$$

$$\sum_{n=0}^{\infty} 2p_n = 2\left(\frac{1}{6} + \frac{1}{3}\right) = 1.$$

A process to generate this distribution using black and white balls analogous to the process described for the distribution in the example of the last section will not have quite the same pleasing simplicity. You need to continue the following random selections until a white ball is drawn. First select a ball randomly from a box of one white and five black balls. If a black ball is selected, select from a box of 2 white and 3 black balls. Then select from a box of 11 white and 13 black balls. For each subsequent selection double the number of white balls and add three more, and double the number of black balls and add one more.

$$2^2p_2 = 2^1p_1 + 2^1\left(\frac{1}{4}\right)^2 = 2^1p_1 + \frac{1}{2}\left(\frac{1}{2}\right)^2,$$

and in general, when $k \geq 1$,

$$2^kp_k = 2^{k-1}p_{k-1} + 2^{k-1}\left(\frac{1}{4}\right)^k = 2^{k-1}p_{k-1} + \frac{1}{2}\left(\frac{1}{2}\right)^k.$$

The first term of the sequence, then, is $^1/_{12}$, and each subsequent term, s_k, is equal to its predecessor, s_{k-1}, plus $^1/_2(^1/_2)^k$, that is, s_{k-1} plus the k^{th} value in the sequence $^1/_4$, $^1/_8$, $^1/_{16}$, $^1/_{32}$, ..., the members of which sum to $^1/_2$. Thus the sequence increases steadily, converging to the limit $^1/_{12} + \sum_{k=1}^{\infty} {}^1/_2(^1/_2)^k = {}^7/_{12}$. Cashed out, the sequence goes

$$\frac{1}{12}, \frac{1}{3}, \frac{11}{24}, \frac{25}{48}, \frac{53}{96}, \frac{109}{192}, \frac{221}{384}, \frac{445}{768}, \frac{893}{1536}, \frac{1789}{3072}, \ldots$$

approaching $^7/_{12}$ ($= {}^{1792}/_{3072}$) asymptotically. So $E(E(B|A) - A) = {}^7/_{12}$.

Obviously the expected gain is always positive, since, for every n, $p_n > {}^1/_2\, p_{n+1}$. Even those who are sceptical about infinite expectations must find this example paradoxical, since the average expected gain conditional on A is finite.[10]

The infinite series

$$E(E(B|A) - A) = 2^0p_0 + \sum_{n=1}^{\infty}(-2^{n-1} + 2^np_n)$$

begins

$$\frac{1}{12} + \left(-\frac{1}{12} + \frac{1}{3}\right) + \left(-\frac{1}{3} + \frac{11}{24}\right) + \left(-\frac{11}{24} + \frac{25}{48}\right) + \ldots$$

It cannot be rebracketed as

$$\left(\frac{1}{12} - \frac{1}{12}\right) + \left(\frac{1}{3} - \frac{1}{3}\right) + \left(\frac{11}{24} - \frac{11}{24}\right) + \ldots$$

because then its sum would be 0, whereas in the original arrangement the sum is $^7/_{12}$, and a convergent series has a unique sum. Rearranged in this way, it is the series for $E(E(B–A)|A+B))$, the average of the expected gains given the total sum in the two envelopes, which accordingly is unequal to $E(E(B|A) – A)$ in this case. Indeed it is well known that such series can be illicitly rearranged to converge to any real number (see Haggarty 1993, pp. 115–16).

[10] Discrete paradoxical infinite distributions like that given by Broome (1995, pp. 6–7) do not yield a finite value for this expectation. Broome's example is:

$2p_n = {}^1/_3(^2/_3)^n$, $n \geq 0$.

Each partial sum includes the gain on swapping for the outcome $(2^{k-1}, 2^k)$ but not the loss for $(2^k, 2^{k-1})$. For example, when $k = 4$, the sum includes the gain for (8, 16) but not the loss for (16, 8).

It is easy to verify that the expectation for the random variable A, the value in the selected envelope, is not finite. We have seen that the sequence of partial sums $\langle 2^k p_k \rangle$ is not null but tends to the limit $^7/_{12}$, whence the series $\sum_{n=1}^{\infty} 2^n p_n$ must grow indefinitely (diverge to $+\infty$, by the vanishing condition), and so therefore must $E(A) = 2^0 p_0 + \sum_{n=1}^{\infty} (2^n p_{n-1} + 2^n p_n)$ (by the first comparison test, Haggarty 1993, p. 108).

On the other hand, if we consider the average expected gain given the total of the sums in the envelopes, $E(E(B–A|A+B))$, this sums to zero not only for finite probability distributions but also for all infinite ones too:

Finite distribution $\sum_{n=0}^{k} (p_n(2^{n+1} - 2^n) + p_n(2^n - 2^{n+1})) = 0$

Infinite distribution $\sum_{n=0}^{\infty} (p_n(2^{n+1} - 2^n) + p_n(2^n - 2^{n+1})) = 0$

Since each partial sum equals zero, the sequence of` partial sums is null. So $E(E(B|A) - A) \neq E(E(B - A|A+B))$.

If we consider the average expected loss given the sum in the other envelope, $|E(B - E(A|B))|$, we find that there is an argument, parallel to that for $E(E(B|A) - A)$, that sticking is profitable. The k^{th} partial sum of $E(B - E(A|B))$ is $-2^k p_k$. So $E(B - E(A|B)) = \lim_{k \to \infty} -2^k p_k$, which in the present example is $-^7/_{12}$, a *loss* of $^7/_{12}$. Here, each partial sum includes the loss on swapping for the outcome $(2^k, 2^{k-1})$ but not the gain for $(2^{k-1}, 2^k)$. For example, when $k = 4$, the sum includes the loss for (16, 8) but not the gain for (8, 16). Again, $E(B - E(A|B))$ and $E(E(B - A|A+B))$ are not the same.

To summarise, in the present example:

(1) $E(E(B|A) - A)$ converges to a positive limit;

(2) $E(E(B–A|A+B))$ converges to 0;

(3) $E(B - E(A|B))$ converges to a negative limit.

These are to be distinguished from $E(B–A)$ and $E(A–B)$, which are equivalent to them only when they are all equal to 0.[11] With the distribution used as an example above, the infinite series for the former goes

$$\frac{1}{12} - \frac{1}{12} + \frac{1}{3} - \frac{1}{3} + \frac{11}{24} - \frac{11}{24} + \frac{25}{48} + \ldots$$

and the sequence of its partial sums oscillates between $2^k p_k$ and 0:

$$\frac{1}{12}, 0, \frac{1}{3}, 0, \frac{11}{24}, 0, \frac{25}{48}, \ldots$$

Here the infinite series is divergent by oscillation; and so, by symmetry, is $E(A–B)$. Scott and Scott (1997) and Arntzenius and McCarthy (1997), unlike Broome (1995), both appear to resolve the paradox by construing the paradoxical cases in those terms. But if the average expected gain is calculated in the way it has been above, as $E(E(B|A) – A)$, it *is* defined. Even those who argue that it is undefined in the cases where the average expected gain is not finite will have to admit it is defined in the example used here.

Now, since $E(A)$ and $E(B)$ are both infinite, it may seem that this result does not involve any inconsistency. If the expected value of each envelope is infinite, can there be any difference between those values? If x is finite, certainly $(^7/_{12} + x) > x$, but if it is not we cannot say that—which, in the words of Chalmers (1996), is "just another example of a familiar phenomenon, the strange behaviour of infinity".

However, pointing this out will not resolve the paradox. It does not distinguish the two-envelope case from the following variant, where there is a genuine positive expected gain. Determine A according to the probability distribution we have been using in this section: $p_0 = {}^1/_{12}$, $p_{n+1} = {}^1/_2\, p_n + {}^1/_2({}^1/_4)^n$ for $n \geq 1$. Let the probability of $A = 1$ be p_0 and the probability of $A = 2^n$, $n \geq 1$, be $p_{n-1} + p_n$. Then determine B thus: if $A = 1$, $B = 2$, else B = half or double A, with the probability ratios $p_{n–1} : p_n$. If A comes out as 4, for example, the expectation for B is $(2p_1 + 8p_2)/(p_1 + p_2) = {}^4/_9$. You are handed the first sealed envelope containing the sum A and given the option of swapping it for the second, whose content is B. It will always be rational to swap and the average expected gain will be $^7/_{12,}$ though the expectations for each envelope have no finite mean.

[11] The series for $E(B – A)$ is alternating: $\sum_{n=0}^{\infty} d_n = 2^0 p_0 – 2^0 p_0 + 2^1 p_1 – 2^1 p_1 + 2^2 p_2 – \ldots$. Clearly its partial sums alternate between 0 and $2^n p_n$, so that, whenever the sequence $\langle 2^n p_n \rangle$ is null, $\sum_{n=0}^{\infty} d_n$ will converge to zero, and will be divergent otherwise. For more detail see the Appendix.

In the paradoxical envelope cases only one of the three results on swapping—(1) positive average gain, (2) zero average gain, (3) average loss—can be correct.

Cases (1), (2) and (3) correspond to calculating the expected expected-gain on swapping by $E(E(B|A)-A)$, $E(E(B-A|A+B))$ and $E(B-E(A|B))$ respectively.[12] Each of the series for these expectations can be constructed by various regroupings of all the terms for the series for $E(B-A)$ (as can be seen by inspecting the beginning of the Appendix), which is to say that these series exhibit certain symmetries as completed infinite series. If those symmetries were preserved in the sequences of partial sums by which the sums of those infinite series are defined, the fact that $E(E(B-A|A+B)) = 0$ would allow us to conclude that they all summed to zero. However, those symmetries are absent in the sequences of partial sums, and it is for this reason that we cannot in general assume that these different series will have the same sum. One subtlety of the paradox is that it disguises this from us. A further subtlety is that by appealing to the symmetry in a distorted manner (equiprobability of our envelope being the larger or the smaller used to tempt us to assume equiprobability of B being half or twice A when given $A = 2^n$) it insinuates the presumption that irrespective of the probability distribution concerned we can disregard the symmetry of the setup when choosing how to calculate the expectations. But that symmetry is significant, since it cannot matter which envelope we called A. As we show in the next paragraph, choosing to calculate the average expected gain on swapping by calculating $E(E(B|A) - A)$ or $E(B - E(A|B))$ ignores the symmetry of the setup. Yet any consideration (un)favourable to one is equally (un)favourable to the other, and they cannot both be correct unless they are both zero, since one is the negative of the other. What follows, then, is the reason why $E(E(B–A|A+B))$ is the uniquely correct way to calculate the expected expected-gain in the paradoxical cases.

What distinguishes $E(E(B–A|A+B))$ as correct is that only in this case do the partial sums of the expectations properly respect the set-up, namely that, *wherever* you can have 2^n in your envelope and twice as much in the other, you can, with equal probability, have 2^{n+1} in yours and half as much in the other, and vice versa. It does so because each of its terms, given by $E(B–A|A+B{=}3.2^n)$, consists in the case where you have 2^n and the other has twice as much, and you have 2^{n+1} and the other has half as much. In the case of $E(E(B|A) - A)$, while the possibility that 2^{n+1} is in the other envelope is included in the partial sums for this expectation, the case in which 2^{n+1} is in your own is excluded; and vice versa for $E(B - E(A|B))$.

[12] In the Appendix we give mathematical reasons for dismissing $E(B - A)$ and $E(B) - E(A)$ as ways of calculating this expectation.

So the partial sums of $E(E(B|A) - A)$ always include a gain without its symmetrical loss, while the partial sums of $E(B - E(A|B))$ always include a loss without its symmetrical gain.[13] Thus only $E(E(B - A)|A + B)$, which gives zero gain, respects the symmetry of the set-up for every finite partial sum. But it is precisely in terms of the sequence of partial sums that the average expected gain is defined.

It may seem that these considerations are not enough to distinguish the standard two-envelope case from the variant described above in which the sum in the second envelope is always determined by a separate procedure as half or double the first. If there is no upper bound on the sums involved in either case, what is the difference? In the variant case must not an outcome with 2^{n+1} in the first envelope and 2^n in the second be possible whenever $(2^n, 2^{n+1})$ is possible (even if they are not equiprobable)? But if we see the infinite cases as extrapolations from or extensions of the finite ones, we can see that partial sums which exclude one of these two possibilities but not the other distort the set-up in the standard case but not in the variant one.

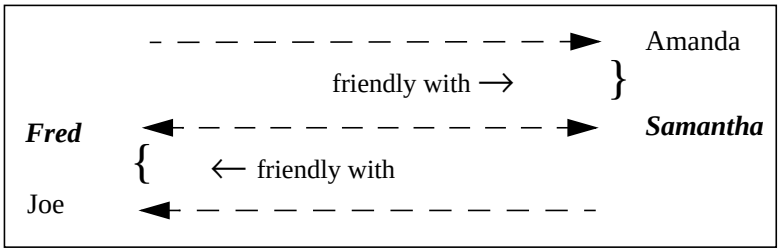
It now emerges that it was a mistake to assimilate the two-envelope paradox to the St. Petersburg paradox as Broome (1995) and Arntzenius and MacCarthy (1997) do. The St. Petersburg paradox turns on our unwillingness to invest a huge finite sum on the basis of an infinite expectation. The essence of the two-envelope paradox is that there is a way of calculating the mean expected gain on swapping that makes it come out non-zero. As we have seen, the best paradoxical two-envelope cases are those in which, though the expectation for your envelope is not finite, the expectation of $E(E(B|A) - A)$ is clearly well-defined and finite.

Suppose that instead of keeping the contents of the envelope you finally settle for, you get nothing unless you swap, in which case you receive the amount gained or pay the sum lost on swapping. This does not change the paradox significantly, and removes any confusion that might arise from the fact that neither envelope has a finite expectation in the problematic cases. So, by contrast with the St. Petersburg paradox, it is not an infinite expectation which makes the problematic two-envelope cases paradoxical.

[13] In terms of a table like Fig. 1, the partial sums for $E(E(B|A) - A)$ end on a complete row, thereby missing out $-2^n p_n$; and the partial sums for $E(B - E(A|B))$ end on a complete column thereby missing out $2^n p_n$. The partial sums for $E(E(B–A|A+B))$, of course, step down the diagonals.

4. Looking inside your envelope

At first sight the paradox seems to return if it is supposed that, before choosing whether to swap, you look inside your envelope and therefore know what sum it contains. Now the expected gain for the paradoxical distributions is unquestionably positive. For example, if value in your envelope is 4, the other envelope must have 2 or 8. Suppose it actually has 2.[14] If you had looked inside the other envelope instead, you would have known the first envelope contained 1 or 4, and then your expected gain on swapping would be negative. But then your knowledge would be different in the two cases. In one case you know the value of *A*, in the other of *B*. Their evidential value on their own is different. For an analogy, suppose my wife and I know that Fred is friendly with Amanda and Samantha, and that Samantha is friendly with Fred and Joe. We catch sight of a couple arriving at a party, but in the darkness I recognize Fred and do not see his companion. My wife, who is some distance in front of me, spots Samantha but does not see who is with her. I conclude that the Fred is with Amanda or Samantha, she concludes that Samantha is with Fred or Joe.



We both have partial, but different evidence, on the basis of which we reach our different conclusions. Nothing odd about this.

Knowing the contents of your envelope before choosing whether to swap does nevertheless give you an advantage, since you will always gain by swapping if you find 1 in your envelope.

This probably does not remove the puzzlement, since it seems that swapping *every* time after looking is better "on average". But what is to be meant by "on average"? It might seem that we are talking about a situation in which you have different information so that the considerations of §3 do not apply. But that is incorrect. Although if you looked in the enve-

[14] Indeed, this apparent resurrection of the paradox does not depend on your actually opening your envelope, since even when your envelope remains sealed you nevertheless know this *would* be the situation if you had opened it.

lope you would have different information, whether swapping is better on average is unaffected by that knowledge. So the considerations of §3 still apply. Rather, the effect of talking about looking in your envelope is to persuade us to return to calculating $E(B|A) - A$, and we know $E(B|A) - A$ is positive for any particular value of A since we have chosen paradoxical distributions that make it so. Now, with the thought that it is better "on average", it seems that we should calculate the average of $E(B|A) - A$, that is, $E(E(B|A) - A)$. Thus are we beguiled once more into calculating the wrong expectation, wrong for the reasons already given in §3.

To elaborate, let us suppose that you take part in thousands of repeated trials with possibly different total sums, with the probabilities given above (which you know). You are allowed to see what is in your envelope before deciding whether to swap but you have to pay a small premium, say a quarter of the expected gain, to swap, except that if you have 1 in your envelope the premium is 1. A general policy of swapping is likely to lose you money, and the more trials there are the more likely this is.

To see this, it may help to suppose that I get the other envelope each time, and look in it, without of course knowing what is in yours. I am allowed to keep a sum equal to what I find or choose to receive a sum which is the same as that in your envelope, but I must pay a premium of a quarter of *my* expected gain. If we both pay a premium each time we cannot both profit from the exchange, and at least one of us must lose if we always choose the other sum: the more trials there are the more likely we both lose.

If all those cases where you have 2 in your envelope are picked out then the average gain for them is likely to be positive. In considering the average gain for a given value in your envelope we are not considering a truly representative sample, one for which we are as likely to have the larger sum in our envelope as the smaller. The calculated expected gain when you find 2 in your envelope takes account only of the outcomes (2, 1) and (2, 4), not of the respectively equiprobable (1, 2) and (4, 2). Now, in general, the average gain when you have 2^n in your envelope is likely to be positive. And, if every possible value in your envelope is allowed for, is not every possible outcome embraced? Yes, but surprisingly it does not follow that your average gain over the whole run is likely to be positive. For, as we have already shown, in the paradoxical cases $E(E(B|A) - A) \neq E(E(B - A|A+B))$. Finally, the argument that, since you would want to swap if you looked you ought to swap anyway (see footnote 13), arbitrarily excludes what you would want to do if you looked in the other envelope, since if you looked in the other envelope you would not want to swap.

Consequently, it would be rational to ignore what you know about the contents of your envelope.

It is as if you were trying to predict the verdicts in criminal trials, knowing that defendants are usually found guilty but being acquainted only with the defence case in each instance. On the basis of this selective information it might well be that the probability of a finding of guilt would generally be lower than that of acquittal, but you are likely to be more accurate in your predictions if you ignore this information and generally predict a guilty verdict.

5. Appendix

5.1. Discrete cases

For discrete infinite probability distributions there are five different series which may be held to give the average expected gain on swapping.

$$E(E(B|A) - A) = \sum_{n=0}^{\infty} a_n,$$

where $a_0 = 2^0 p_0$ and $a_n = 2^n p_n - 2^{n-1} p_{n-1}$ for $n > 0$

$$= (2^0 p_0) + (-2^0 p_0 + 2^1 p_1) + (-2^1 p_1 + 2^2 p_2) + \ldots \qquad (1)$$

$$E(B - E(A|B)) = \sum_{n=0}^{\infty} b_n,$$

where $b_0 = -2^0 p_0$ and $b_n = 2^{n-1} p_{n-1} - 2^n p_n$ for $n > 0$

$$= (-2^0 p_0) + (2^0 p_0 - 2^1 p_1) + (2^1 p_1 - 2^2 p_2) + \ldots \qquad (2)$$

$$E(E(B-A|A+B)) = \sum_{n=0}^{\infty} c_n, \text{ where } c_n = 2^n p_n - 2^n p_n$$

$$= (2^0 p_0 - 2^0 p_0) + (2^1 p_1 - 2p^1{}_1) + (2^2 p_2 - 2^2 p_2) + \qquad (3)$$

$$E(B-A) = \sum_{n=0}^{\infty} d_n, \text{ where } d_n = \begin{cases} 2^m p_m \text{ if } n = 2m, \text{ i.e. } n \text{ even} \\ -2^m p_m \text{ if } n = 2m+1, \text{ i.e. } n \text{ odd} \end{cases}$$

$$= (2^0 p_0) + (-2^0 p_0) + (2^1 p_1) + (-2^1 p_1) + (2^2 p_2) + (-2^2 p_2) + \ldots (4)$$

$$E(B) - E(A) = \sum_{n=0}^{\infty} e_n - \sum_{n=0}^{\infty} e_n, \text{ where } e_n = 3.2^n p_n$$

$$\text{(because } (2^0 p_0) + (2^1 p_0 + 2^1 p_1) + \ldots = \sum_{n=0}^{\infty} 3.2^n p_n)$$

$$= \sum_{n=0}^{\infty} 3.2^n p_n - \sum_{n=0}^{\infty} 3.2^n p_n$$

$$= ((3.2^0 p_0) + (3.2^1 p_1) + (3.2^2 p_2) + \ldots)$$

$$- ((3.2^0 p_0) + (3.2^1 p_1) + (3.2^2 p_2) + \ldots) \qquad (5)$$

Brackets have been used to emphasise individual terms of the series.

That these are all distinct series is evident from the fact that they are defined by sequences which are not identical: $\langle a_n \rangle \neq \langle b_n \rangle \neq \langle c_n \rangle \neq \langle d_n \rangle$. (5) is the difference between two series, each of which is $\sum e_n$. Of course, they may have the same sum, but that has to be proved, and can be proved only by considering the sequences of their partial sums, which in turn can only be determined from the sequences which define their terms.

The sum of $\sum c_n$ is always defined, and is always zero.

If the alternating series $\sum d_n$ has a sum, it sums to zero, but unless the sequence of its partial sums,

$$\langle x_n \rangle \text{ where } x_n = \begin{cases} 2^m p_m \text{ if } n = 2m, \text{ i.e. } n \text{ even} \\ 0 \text{ if } n = 2m+1, \text{ i.e. } n \text{ odd} \end{cases}$$

is null, which as noted above in footnote 11 will be the case when $\langle 2^n p_n \rangle$ is null, it will be divergent and so will not sum to anything. Arbitrary regroupings of all the terms from an alternating series such as $\sum d_n$ will give series with the same sums only if the alternating series is *absolutely convergent*, that is, in this case, iff $\sum |d_n|$ is convergent.

When $\sum d_n$ is not absolutely convergent, then, as we noted in §3 above, for any real number there will be a way of grouping all the members of $\langle d_n \rangle$ into a new series that converges to that number. If an alternating infinite series which is not absolutely convergent were equivalent to any regrouping of it, then any number could be proved equal to any other number. Notoriously, we would have

$$\begin{aligned} 0 &= 0 + 0 + 0 + ... \\ &= (1 - 1) + (1 - 1) + (1 - 1) + ... \\ &= 1 - 1 + 1 - 1 + 1 - 1 + ... \\ &= 1 + (-1 + 1) + (1 - 1) + (1 - 1) + ... \\ &= 1 + 0 + 0 + 0 + ... \\ &= 1. \end{aligned}$$

LEMMA *E*(*A*) *is finite if and only if* $\sum d_n$ *is absolutely convergent*.

PROOF

Only if. If $E(A) = \sum_{n=0}^{\infty} 3.2^n p_n$, is finite, and so convergent, then $\sum_{n=0}^{\infty} 2^n p_n$ is convergent by the scalar product rule (Haggarty 1993, p. 106).

The n^{th} partial sums of $\sum |d_n|$ are

$$\begin{cases} 2 \sum_{k=0}^{m} 2^k p_k \text{ if } n = 2m+1 \\ 2 \sum_{k=0}^{m-1} 2^k p_k + 2^m p_m \text{ if } n = 2m \end{cases}$$

So the sequence of partial sums $\Sigma|d_n|$ has an odd and an even subsequence (ibid., p. 88). So long as they converge to the same limit then Σd_n will converge to that limit. The odd subsequence is $\langle 2\sum_{k=0}^{m} 2^k p_k\rangle$ and, since $\sum_{n=0}^{\infty} 2^n p_n$ is convergent and equals $^1/_3E(A)$, $\lim_{m\to\infty}\langle 2\sum_{k=0}^{m} 2^k p_k\rangle = {}^2{}_3E(A)$. The even subsequence is

$\langle 2\sum_{k=0}^{m-1} 2^k p_k + 2^m p_m\rangle$. Since $\sum_{n=0}^{\infty} 2^n p_n$ is convergent, we know that $\langle 2^m p_m\rangle$ is null. Hence the even subsequence is a sum of convergent sequences, so that

$$\begin{aligned}&\lim_{m\to\infty}\langle 2\sum_{k=0}^{m-1} 2^k p_k + 2^m p_m\rangle\\ &= \lim_{m\to\infty}\langle 2\sum_{k=0}^{m-1} 2^k p_k\rangle + \lim_{m\to\infty}\langle 2^m p_m\rangle\\ &= \tfrac{2}{3}E(A) + 0 = \tfrac{2}{3}E(A) \text{ (ibid., p. 73, sum rule).}\end{aligned}$$

Therefore $\Sigma|d_n|$ converges, so that Σd_n is absolutely convergent.

If. Suppose that Σd_n is absolutely convergent. Then $\Sigma|d_n|$ is absolutely convergent. $\Sigma|d_n| = (2^0p_0) + (2^0p_0) + (2^1p_1) + (2^1p_1) + \ldots = (2^0p_0 + 2^0p_0) + (2^1p_1 + 2^1p_1) + \ldots$ by the rearrangement rule (ibid., p. 116), $= \sum_{n=0}^{\infty} 2.2^n p_n$. So $\sum_{n=0}^{\infty} 2.2^n p_n$ is convergent, whence by the scalar product rule $E(A) = \sum_{n=0}^{\infty} 3.2^n p_n$ is convergent. Thus $E(A)$ is finite.

(At this point it is worth recalling that in speaking of the four series, Σa_n to Σd_n, we are speaking of the behaviour of the corresponding expectations, and that paradox appears when these series and (5) do not have the same sum.)

So, by this lemma:

(a) whenever $E(A)$ is finite, $\sum d_n$ is absolutely convergent and hence series (1) to (5) are all equal by the rearrangement rule, since they are all composed of rearrangements of $\sum d_n$;

(b) whenever $E(A)$ is not finite, $\sum d_n$ is not absolutely convergent, and we cannot appeal to the rearrangement rule to prove (1) to (5) equal.

But (b) does not mean that (1) to (5) must all be divergent, or that some of them may not be equal. In this case we must assess their sums by considering the sequences of their partial sums, which turn out to depend on the sequence $\langle 2^n p_n \rangle$, the sequence of partial sums of $\sum a_n$, that is, of $E(E(B|A) - A)$, which as we know is closely related to $E(A)$.

So when $E(A)$ is finite no paradoxes arise. When $E(A)$ is not finite we can have both non-paradoxical and paradoxical cases.

We present the following results:

When $E(A)$ is not finite, the sequence $\langle 2^n p_n \rangle$ may

(i) be null. In this case $\sum a_n = \sum b_n = \sum c_n = \sum d_n = 0$, and (5) is undefined (being the difference of two divergent sequences). An example of a probability distribution which has this effect is:

$$2p_0 = 1 - \sum_{n=1}^{\infty} 2p_n \text{; and, for n} \geq 1,\ 2p_n = \frac{1}{n+1}\left(\frac{1}{2}\right)^{n-1}.$$

(ii) converge to a finite positive number x, so that $\sum a_n > 0$, $\sum b_n < 0$ (because if $\sum b_n$ has a sum it will be the limit of the sequence of negative terms, $\langle -2^n p_n \rangle$), $\sum c_n = 0$, $\sum d_n$ oscillates and (5) is undefined. In this case the average expected gain given by $\sum a_n$ may (as in the example of §3) or may not involve an expected gain for *every* value of A. An example of a probability distribution for the latter case is:

$$2p_0 = \frac{1}{2};\ 2p_n = p_{n-1},\ n \geq 1.$$

Here the average expected gain is $^1/_4$; the only value of A for which there is a gain is 1, and for $A > 1$ the expected gain is 0.

(iii) diverge to $+\infty$. See footnote 10 for an example in the literature. $\sum a_n$ diverges to $+\infty$, $\sum b_n$ diverges to $-\infty$, $\sum c_n = 0$. As with (ii), $\sum d_n$ oscillates and (5) is undefined.

(iv) oscillate. In this case $\sum a_n$, $\sum b_n$ and $\sum d_n$ oscillate, $\sum c_n = 0$, and (5) is undefined.

The following considerations apply:

(a) Expectations for which the corresponding series has no defined sum are themselves undefined and need not be considered.

(b) Expectations for which the series is divergent by oscillation cannot in any sense be said to have a sum, and so are undefined.
(c) Expectations for which the series is divergent to positive or negative infinity may, controversially, be said to have an infinite sum, and as such may be said to be controversially well defined.
(d) Expectations for which the series is convergent are uncontroversially said to have a sum and so may be said to be uncontroversially well defined.

Whenever the expectations are well defined but with different sums we have a potentially paradoxical case. Hence:

1. No paradoxes arise under case (i) since all four of the expectations corresponding to $\sum a_n$, $\sum b_n$, $\sum c_n$ and $\sum d_n$ well defined and equal, and by (a) the other may be disregarded.
2. In case (iv) by (a) and (b) only $E(E(B–A)|A+B)$ is well defined, so no paradox arises.
3. In cases (ii) and (iii) $E(B–A)$ and $E(B) – E(A)$ are undefined by (b) and (a) respectively, so need not be considered.
4. In case (ii) all three of the expectations corresponding to $\sum a_n$, $\sum b_n$, and $\sum c_n$ are uncontroversially well defined by (d), and being unequal give rise to the sort of paradox not previously discussed in the literature, which we refer to as "best paradoxical".
5. In case (iii) the expectations corresponding to $\sum a_n$, and $\sum b_n$ are controversially well defined by (c)—with $\sum c_n = 0$ as always—and being unequal give rise to the sort of paradox previously discussed in the literature, which we refer to as "unbounded paradoxical".

For these reasons in paradoxical cases (cases (ii) and (iii)) we need to choose between, and *only* between, the expectations corresponding to $\sum a_n$, $\sum b_n$, and $\sum c_n$.

5.2. Continuous cases

We conclude with a note on continuous distributions, for the version of the paradox in which money is treated as if it were a continuous rather than a discrete quantity.

For continuous as well as discrete cases, whenever the relevant infinite series and improper integrals are well-behaved, which is always the case when $E(A)$ is finite, we can prove using standard probability theorems that

$$\begin{aligned} E(B) – E(A) &= E(B – A) \\ &= E(E(B – A|A + B)) \\ &= E(E(B|A) – A) \end{aligned}$$

$$= - E(A - E(B|A))$$
$$= - E(B - E(A|B)) \text{ by symmetry.}$$

Since $E(A) = E(B)$, each of these = 0. (Cf. Broome 1995, pp. 10–11.) They hold in virtue of

(i) $E(X - Y) = E(X) - E(Y)$ (see, for example, DeGroot 1989, p. 188, theorem 3)

(ii) $E(E(X|Y)) = E(X)$ (ibid., p. 220, theorem 1)

(iii) $E(aX) = aE(X)$ (ibid., p. 187, theorem 1).

Just as in the discrete case, there are probability density functions where the values of the envelopes do not have a finite expectation but $E(E(B|A) - A)$ is finite, and others too where the latter expectation is infinite, as we proceed to demonstrate.

Let the random variable S represent the smaller sum in the two envelopes, and let it have probability density function $f(s)$ and cumulative distribution function $F(s)$, such that S takes values only from $[0, \infty)$, that is, $F(0) = 0$, $F(\infty) = 1$. Let the random variables A and B represent the sums in the envelopes, and their p.d.f.'s and c.d.f.'s be $f_A(a)$, $f_B(b)$, $F_A(a)$, $F_B(b)$. We shall determine these as functions of $f(s)$. Obviously $f_A(a) = f_B(b)$ and $F_A(a) = F_B(b)$. To determine the p.d.f. of A we consider its c.d.f. $F_A(a) = \Pr(A \leq a)$. (See Degroot 1989, pp. 102ff.)

Now $A \leq a$ (i) whenever $0 \leq s \leq a/2$, since, whether A is the smaller sum or the larger, it is still less than or equal to a; or (ii) when $a/2 < s \leq a$ and $A = s$.

So $F_A(a) = \Pr(A \leq a) = \Pr(0 \leq s \leq a/2 \text{ or } (a/2 < s \leq a \text{ and } A = s))$
$$= F(a/2) - F(0) + \frac{1}{2}\Pr(a/2 < s \leq a) \text{ (since } \Pr(A = s) = {}^1/_2)$$
$$= F(a/2) + \frac{1}{2}(F(a) - F(a/2))$$
$$= \frac{1}{2}F(a) + \frac{1}{2}F(a/2).$$

Since $f_A(a) = \dfrac{dF_A}{da}$ we have p.d.f. of A,

$$f_A(a) = \frac{d}{da}(\tfrac{1}{2}F(a) + \tfrac{1}{2}F(a/2))$$
$$= \tfrac{1}{2}f(a) + \tfrac{1}{4}f(a/2).$$

So $E(A)$ tends to infinity if and only if $E(S)$ does.

We have

$$E(B|A) = a\frac{2f(a) + \frac{1}{4}f(a/2)}{f(a) + \frac{1}{2}f(a/2)} \text{ from Broome (1995, p. 10)}^{15}$$

$$= a\frac{2f(a) + \frac{1}{4}f(a/2)}{2f_A(a)}$$

and hence

$$E(E(B|A) - A) = \int_0^\infty (E(B|A) - A) f_A(a)da$$

$$= \int_0^\infty \left(a\frac{2fa) + \frac{1}{4}f(a/2)}{2f_A(a)} - a \right) f_A(a)da$$

$$= \int_0^\infty \frac{af(a)}{2} - \frac{af(a/2)}{8}da.$$

Now if $\int_0^\infty \frac{af(a)}{2}da$ and $\int_0^\infty \frac{af(a/2)}{8}da$ exist, which they do if the expectation of the smaller sum, $E(S)$, is finite, then we can take a further step:

$$E(E(B|A) - A) = \int_0^\infty \frac{af(a)}{2}da - \int_0^\infty \frac{af(a/2)}{8}da = 0.$$

When $E(S)$ is not finite, then since $E(S) = \int_0^\infty sf(s)ds$, which by definition equals $\lim_{x \to \infty} \int_0^x sf(s)ds$, it means that $\lim_{x \to \infty} \int_0^x sf(s)ds$ does not exist, either because $\lim_{x \to \infty} \int_0^x sf(s)ds$ is unbounded as $x \to \infty$ (when we say that $E(S)$ is infinite), or because the integral oscillates as $x \to \infty$. In this case we cannot take the further step and must evaluate

[15] Arntzenius and McCarthy (1997, p. 40) criticise Broome's derivation as inadequate. Nevertheless, the result is correct, and it can be derived in an unimpeachable way by considering $E(B|A{=}a) = \lim_{\delta a \to 0} E(B|a \le A \le a + \delta a)$ if that limit exists, which it does if the c.d.f of the smaller envelope is differentiable.

$$E(E(B|A) - A) = \int_0^\infty \frac{af(a)}{2} - \frac{af(a/2)}{8} da$$

directly.

There is at least one family of p.d.f.'s for S rich in cases where $E(A)$ is infinite (because $E(S)$ is infinite) and yet $E(E(B|A) - A)$ is finite. The family is

$$f(s) = \begin{cases} \dfrac{k}{(as^2 + bs + c)^d} & \text{for } s \in [0, \infty) \\ 0 & \text{for } s \in (-\infty, 0) \end{cases}$$

where k is a normalising constant so that $\int_0^\infty f(s)ds = 1$.

Conjectures:

(a) When $d = 1$ $E(S)$ is infinite yet $E(E(B|A) - A)$ is finite.

(b) When d is in $(^1/_2, 1)$ both $E(S)$ and $E(E(B|A) - A)$ are infinite.

(c) When $d \leq {}^1/_2$ then $\int_0^\infty f(s)ds = \infty$, so that f is not a p.d.f; and when $d > 1$ then $E(S)$ is finite, so that $E(E(B|A) - A) = 0$.

We have a general proof of (a) and proofs of various cases and sub-families of (a), (b) and (c). For example, all three conjectures are true whenever the quadratic in the denominator is a perfect square.

For concreteness, we now give an example where, just as in the discrete case, though the envelopes have no finite expectation there is a perfectly well-defined and finite average expected gain on swapping. For the p.d.f.

$$f(s) = \frac{2}{\pi} \frac{1}{s^2 + 1}$$

we find that $E(A) = E(B) = E(S) = \infty$ while $E(E(B|A) - A) = \log_e 4/(2\pi) \approx {}^2/_9$. The continuous example given by Broome (1995, p. 7), $f(s) = 1/(s + 1)^2$, also turns out to yield a finite average expected gain, since it belongs to the subfamily mentioned at the end of the last paragraph.

We now consider a conjecture which, if proved, gives reason to believe that for continuous cases no extra issues of significance arise. It may be the case that there are pathological p.d.f.'s for which the conjecture would not apply. However, it should apply to all "normal" p.d.f.'s for which the c.d.f. is differentiable. Since at present it is unclear whether any expecta-

tions apart from $E(E(B–A)|A+B))$ are evaluable when the c.d.f. is not differentiable, this may not matter, because then no such pathological cases give rise to paradoxical expectations. However, if the evaluation of other expectations can be got through on a weaker condition, there may yet be continuous cases beyond our present considerations. If anyone can construct such a case and justify the evaluation of differing average expected gains we would be fascinated to see it.

The discrete cases are determined by the behaviour of the sequence $\langle 2^n p_n \rangle$, which is to say, just by $E(E(B|A) - A)$, since $\langle 2^n p_n \rangle$ is the sequence of partial sums of $E(E(B|A) - A)$. The distinctions made among the various average expected gains for the discrete cases made at the beginning of this Appendix could have been phrased in terms of integrals to give a similar range of infinite integrals for the various average expected gains in the continuous cases. They too depend on the behaviour of $E(E(B|A) - A)$. So to show that no extra issues arise it will suffice to show that $E(E(B|A) - A)$ in the discrete case is convergent if and only if some corresponding continuous $E(E(B|A) - A)$ is finite. For then any continuous case will have a corresponding discrete case to which our argument will apply, which in turn will inform us about the continuous case (hence the need for the biconditional).

In the discrete case $E(E(B|A) - A) = \sum_{n=0}^{\infty} a_n$ as defined above. In the continuous case let $h(a) = a\left(\frac{f(a)}{2} - \frac{f(a/2)}{8}\right)$, when $E(E(B|A) - A) = \int_0^{\infty} h(a)da$. Given the restriction that $h(a)$ is continuous, positive and decreasing on $[0, \infty)$, a simplification of the Euler–Maclaurin summation series tells us that $\sum_{n=0}^{\infty} h(n)$ converges if and only if $\int_0^{\infty} h(a)da$ converges (is finite). (Binmore 1982, p. 136; Haggarty 1993, pp. 113, 233–4. This is the integral test for convergence.)

CONJECTURE *$E(E(B|A) - A)$ in the discrete case is convergent if and only if some corresponding continuous $E(E(B|A) - A)$ is finite.*

SKETCH OF PROOF

Only if. We define $\int_0^{\infty} h(a)da$ such that $h(n)$ equals a suitable function of a_n. Provided this function is linear, then the combination rules for integrals together with Euler–Maclaurin give us that $E(E(B|A) - A) =$

$\int_0^\infty h(a)da$ is convergent. Detail is needed here to prove the uniqueness of $h(s)$ and that $f(s)$ derived in the obvious way from $h(s)$ is a p.d.f.

If. We define $\sum_{n=0}^{\infty} a_n$ such that for all n, a_n equals a suitable function of $h(n)$. So long as this function is linear then the combination rules for series together with Euler–Maclaurin give us that $E(E(B|A) — A) = \sum_{n=0}^{\infty} a_n$ is convergent. Detail is needed here to get from $\sum_{n=0}^{\infty} h_n$ to $\sum_{n=0}^{\infty} a_n$ and then to prove that deriving the p_n in the obvious way from the a_n gives us a probability distribution.

We emphasise that we have not carried out this proof. Nevertheless, the conjecture is plausible and the proof sketch looks feasible.

There is nothing about this conjecture which means that what it sets up as corresponding discrete and continuous cases are behaviourally the same qua paradox. The behavioural correspondences it gives are only that:

1. If one converges and is non-paradoxical, the other may be either non-paradoxical or best paradoxical ($E(E(B|A) – A)$ finite and non-zero) but not unbounded paradoxical ($E(E(B|A) – A)$ infinite).
2. If one converges and is paradoxical the other may either be non-paradoxical or best paradoxical but not unbounded paradoxical.
3. If one diverges whether by being infinite or by oscillation the other may diverge to infinity or by oscillation.

The significance of the plausibility of this conjecture is that it gives weight to the assertion that the continuous cases parallel the discrete ones. We believe it is sufficient to claim that the considerations that apply to the various ways of calculating the average expected gains in the discrete cases carry over into the continuous cases. In the body of the paper we argued that $E(E(B–A)|A+B))$ is the correct way to calculate the average expected gain. So for continuous cases we find that since

$$E(B-A|A+B=c) = \frac{1}{2}\left(\frac{2c}{3}-\frac{c}{3}\right)+\frac{1}{2}\left(\frac{c}{3}-\frac{2c}{3}\right) = 0$$

we have

$$E(E(B-A|A+B=c)) = \int_0^\infty E(B-A|A+B=c)\, f_{A+B}(c)dc$$

$$= \int_0^\infty 0 \times f_{A+B}(c)dc = 0.$$

Department of Philosophy MICHAEL CLARK
University of Nottingham
University Park
Nottingham NG7 2RD, UK
michael.clark@nottingham.ac.uk

Department of Mathematics NICHOLAS SHACKEL
De Montfort University
The Gateway, Leicester LE1 9BH, UK
nshackel@dmu.ac.uk

REFERENCES

Arntzenius, Frank and McCarthy, David 1997: "The Two Envelope Paradox and Infinite Expectations", *Analysis* 57, pp. 42–50.

Binmore, K. G, 1982: *Mathematical Analysis: a straightforward approach*, 2nd ed., Cambridge: Cambridge University Press.

Broome, John 1995: "The Two-envelope Paradox", *Analysis* 55, pp. 6–11.

Castel, Paul and Batens, Diderik 1994: "The Two-envelope Paradox: The Infinite Case", *Analysis* 54, 46-49.

Chalmers, David 1996: "The Two-envelope Paradox: A Complete Analysis?", published on the Internet at http://avocado.wustl.edu/~chalmers/chalmers.envelope.html.

DeGroot, Morris H. 1989: *Probability and Statistics*. 2nd ed., Reading, Massachusetts: Addison-Wesley.

Haggarty, Rod 1993: *Fundamentals of Mathematical Analysis*. 2nd ed., Harlow: Addison-Wesley.

Jackson, Frank, Menzies, Peter, and Oppy, Graham: 1994. "The Two Envelope 'Paradox'", *Analysis* 54, pp. 43–45.

Kraitchik, Maurice 1943: *Mathematical Recreations*. London: Allen & Unwin.

Scott Alexander D., and Scott, Michael 1997: "What's in the Two Envelope Paradox?", *Analysis* 57, pp. 34–41.