WHEN THE SUM OF OUR EXPECTATIONS FAILS US: THE EXCHANGE PARADOX

BY

JOHN D. NORTON[1]

Abstract: In the exchange paradox, two players receive envelopes containing different amounts of money. The assignment of the amounts ensures each player has the same probability of receiving each possible amount. Nonetheless, for each specific amount a player may find in his envelope, there is a positive expectation of gain if the player swaps envelopes with the other player, in apparent contradiction with the symmetry of the game. I consider a variant form of the paradox that avoids problems with improper probabilities and I argue that in it these expectations give no grounds for a decision to swap since that decision must be based on a summation of all the expectations. But this sum yields a non-convergent series that has no meaningful value. The conflicting recommendations – that it is to one or the other player's advantage to swap – arise from different ways of grouping terms in the sum that yield an illusion of convergence. I describe a generalized exchange paradox, explore some of its properties and display another example.

1. Introduction

In the exchange paradox, a game is played in which a randomly chosen amount of money is placed in one envelope and twice that amount in a second. The envelopes are then shuffled and randomly assigned to two players such that each player has an equal chance of receiving the first envelope. The players are then given the option of swapping. Player 1 reasons:

Pacific Philosophical Quarterly 79 (1998) 34–58 0031–5621/98/0100–0000
© 1998 University of Southern California and Blackwell Publishers Ltd. Published by Blackwell Publishers, 108 Cowley Road, Oxford OX4 1JF, UK and 350 Main Street, Malden, MA 02148, USA.

My envelope will contain some amount of money – x dollars, say. There is a probability 1/2 that player 2's envelope has 2x and probability 1/2 that it has x/2. Therefore my expectation in swapping is

$$\text{Expectation} = (1/2)(2x - x) + (1/2)(x/2 - x) = x/4 > 0 \qquad (1a)$$

So, no matter what amount x I find in my envelope, there is a positive expectation in swapping. So I should swap.

Player 2 reasons identically and also sees an advantage in swapping. The paradox resides in the symmetry of the game. Neither player should expect any advantage in swapping since each has been assigned an envelope by an identical, random procedure and had an equal chance of receiving either envelope. Moreover the game is a zero sum game. Any action that profits player 1 must result in a loss for player 2 and *vice versa*. So both players cannot expect an advantage in swapping.

This is the original version of the paradox and has been discussed extensively (for example, see Zabell (1988), Nalebuff (1989), Christensen and Utts (1992), Linzer (1994), Sobel (1994) and Rawling (1994)). The essence of the paradox is a tension between two apparently incompatible claims:

(i) For each fixed amount that player 1 may find in his envelope, there is a positive expectation of gain in swapping envelopes.
(ii) The procedure for assigning amounts in the envelopes is symmetrical so there can be no systematic advantage to either player in swapping.

Unfortunately this tension is masked in the original version by a technical difficulty. A probability distribution for the amounts of money in the envelopes cannot be normalized to unity if it is to satisfy the conditions specified. One possibility is to consider non-standard mathematical representations of the distribution, such a mentioned by Sobel (1994, p. 73) and Rawling (1994, p. 98). Instead, in this paper, following Nalebuff (1989, pp. 177–80) and Linzer (1994, p. 418), I will consider a modified version of the paradox that retains the essential tension between (i) and (ii) but allows standard probability distributions. This modified version will be described in section 2.

In section 3, I will argue that claim (i) does hold. But it does not contradict the symmetry of the game claimed in (ii). That symmetry is violated by our stipulation that player 1 has some fixed amount – \$16, say. But the amount in player 2's envelope must be either \$8 or \$32 with various probabilities. So the computation of each player's expectation in swapping derives from asymmetric information and need not give the same value for each player.

© 1998 University of Southern California and Blackwell Publishers Ltd.

In section 4, I will argue that claim (ii) holds as well and that it is compatible with claim (i). Here I follow Nalebuff (1989, pp. 177–80) and Linzer (1994, p. 418) and urge that this compatibility rests on the fact that player 1's expectation of the contents of his envelope diverge. What I add to this observation is a detailed reconstruction of how this divergence allows the paradox to be constructed. Player 1's expectation in swapping is expressed as the sum of infinitely many probabilistically weighted conditional expectations, each derived from assuming some specific amount in each envelope. We will see that this series is non-convergent. Therefore it cannot be said to have a sum so that the expectation has no meaningful value. Nonetheless, since the series has alternating signs, there are ways of grouping its terms that give the illusion that the series has some definite sum. One such grouping derives from the conditional expectation computed as (1a) and seems to reveal a favorable expectation in swapping for player 1. The argument that player 1 would gain in swapping since expectation (1a) is positive for any amount in his envelope is merely an indirect way of effecting summation through this grouping. Therefore it is fallacious. Another grouping fallaciously suggests that it is to player 2's advantage to swap. A third reflects the symmetry of the game and yields an expectation of zero. All these sums are illusions since the expectation has no value and the arguments corresponding to them fallacies.

In section 5, I will reinforce the compatibility of claims (i) and (ii) by trying to exploit (i) to devise a winning strategy for player 1. We shall see that the strategies proposed succeed only when they violate the symmetrical conditions of play presumed in (ii). In section 6, I will set up a general characterization of exchange type paradoxes that allows us to create as many more paradoxes as we have patience and ingenuity at story-telling. Within this framework, I will prove a generalized version of Nalebuff's (1989, p. 179) result that the exchange paradox is inevitably associated with divergent expectations. Finally, in section 7, I will display another example of an exchange paradox.

2. The Modified Version of the Paradox

To treat the paradox precisely, we need to know more about how the various amounts that may be in the envelopes are chosen. The original statement of the paradox is vague here. Let

(x,y)

represent the state in which the random procedure assigns x dollars to player 1's envelope ("envelope 1") and y dollars to player 2's envelope

© 1998 University of Southern California and Blackwell Publishers Ltd.

("envelope 2"). What are the probabilities of each of the possible states? What are the possible states? To proceed we must fill in the details in some way. I propose to do so in a way that preserves the essential claims (i) and (ii) but otherwise keeps the problem as simple as possible. Consistent with this requirement, we can assume that the randomized selection mechanism allows only the states

$$(1,2), (2,1), (2,4), (4,2), (4,8), (8,4), ..., (2^n,2^{n+1}), (2^{n+1},2^n),...$$

Notice that it is essential for the paradox that there be no upper bound on the amount that may be in the envelope. Otherwise, the paradox is trivially escaped. For swapping when envelope 1 contains this maximum amount is an assured loss so that claim (i) no longer holds. We do assume, however, that there is a minimum amount, $1. This actually strengthens the paradox in this sense. If a player has $1 in his envelope, then there is not just a probabilistic expectation of gain in swapping; there is a certainty of gain since the other envelope must contain $2.

We also assume some probability distribution over these states, defining p_n as

$$p_n = P(2^n,2^{n+1}) = P(2^{n+1},2^n) \quad (2)$$

for all n. This equality of probability of $(2^n,2^{n+1})$ and $(2^{n+1},2^n)$ is essential if the symmetry of claim (ii) is to be preserved. The conditions of the paradox would lead us to expect that, if we fix the content of envelope 1 at 2^n, then there should be equal probability that there is either 2^{n+1} or 2^{n-1} in envelope 2. That is

$$p_n = P(2^n,2^{n+1}) = P(2^n,2^{n-1}) = p_{n-1} \quad (3a)$$

But it is easy to see that, if we allow infinitely many states, condition (3a) leads to an improper probability distribution over all the states, for we have

$$p_0 = p_1 = p_2 = p_3 = p_4 = ... = p_n = ...$$

We could try to accommodate (3a) by invoking non-standard mathematics, such as mentioned by Sobel (1994, p.73) and Rawling (1994, p. 98). A far simpler approach is to modify the statement of the paradox so that its essential character is retained, but (3a) is replaced by another relation that does not lead to improper probabilities. We replace (3a) by

$$p_n = P(2^n,2^{n+1}) = kP(2^n,2^{n-1}) = kp_{n-1} \quad (3b)$$

© 1998 University of Southern California and Blackwell Publishers Ltd.

where we set

$$1/2 < k < 1 \tag{4}$$

for reasons that will be apparent shortly. So we must now have p_n reducing geometrically with increasing n for

$$p_n = kp_{n-1} = k^2 p_{n-2} = \ldots = k^n p_0$$

That p_n decrease geometrically and also sum over all states to 1 forces a unique choice[2]

$$\begin{aligned} p_0 &= P(1,2) = P(2,1) = (1/2)(1-k) \\ p_1 &= P(2,4) = P(4,2) = (1/2)(1-k)k^1 \\ p_2 &= P(4,8) = P(8,4) = (1/2)(1-k)k^2 \\ &\ldots \\ p_n &= P(2^n,2^{n+1}) = P(2^{n+1},2^n) = (1/2)(1-k)k^n \\ &\ldots \end{aligned} \tag{5}$$

The most important point is that probability distribution (5) preserves the paradox as long as $1/2<k<1$. For if player 1 finds 2^n (where $n>0$) in his envelope, then the probability that there is 2^{n-1} in the other envelope is

$$\frac{(1/2)(1-k)k^{n-1}}{(1/2)(1-k)k^{n-1} + (1/2)(1-k)k^n} = \frac{1}{1+k}$$

and the probability that there is 2^{n+1} in the other envelope is

$$\frac{(1/2)(1-k)k^{n}}{(1/2)(1-k)k^{n-1} + (1/2)(1-k)k^n} = \frac{k}{1+k}$$

Therefore, in place of (1a) in the original formulation of the paradox, we have his expectation (for $n>0$) in swapping as

$$\text{Expectation} = \frac{k}{1+k}(2^{n+1} - 2^n) + \frac{1}{1+k}(2^{n-1} - 2^n) = 2^{n-1}\frac{2k-1}{1+k}$$

This expectation is greater than zero when $1/2<k$. We rule out $k\geq 1$ since then the probabilities would become improper.[3] Thus condition (4) gives us well behaved probabilities and preserves the paradox in so far as player 1 retains a positive expectation in swapping for any specific amount that he may find in his envelope.

© 1998 University of Southern California and Blackwell Publishers Ltd.

It will be convenient to re-express this expectation as a conditional expectation. Let the random variables e_1 and e_2 be the amounts in the envelopes of players 1 and 2 prior to their decision to swap. If they swap, then player 1 gains e_2–e_1. His expectation, conditioned on his having 2^n in his envelope, is written as

$$\begin{aligned} E(e_2 - e_1 | e_1 = 2^n) &= 2^{n-1}\frac{2k-1}{1+k} \quad \text{for } n>0 \\ &= 1 \quad \text{for } n = 0 \end{aligned} \tag{1b}$$

3. *Defense of Claim (i): Why Player 1 has an Expectation of Gain from a Swap if the Amount in his Envelope is Known*

The argument is simple and inescapable. We will assume that player 1's envelope has some fixed amount in it – say \$16 for concreteness – although nothing in the following will depend on the particular value chosen. Thus the possible states of the world are

(\$16, \$8) and (\$16, \$32) (6a)

and their probabilities are

$$P(\$16,\$32) = \frac{k}{1+k} \qquad P(\$16,\$8) = \frac{1}{1+k} \tag{6b}$$

We can represent the information needed to decide player 1's expectation in a swap by means of a payoff matrix. The columns correspond to states, the rows to the envelopes; and each cell indicates the amount in the given envelope in the given state.

	(\$16, \$8)	(\$16, \$32)
Envelope 1	\$16	\$16
Envelope 2	\$8	\$32

Matrix 1: Payoffs when player 1's envelope has \$100

Given the probabilities (6b) of the two states of the world, the expectation for player 1 in swapping to envelope 2 is calculated by (1b) and is $8\frac{2k-1}{1+k} > 0$.

© 1998 University of Southern California and Blackwell Publishers Ltd.

The problem with this calculation is that it seems to violate the symmetry of the game. But this is only an illusion. What violates the symmetry is our *stipulation* in (6a) of the two possible states. This treats player 1 and 2 asymmetrically. Therefore any calculation based on this stipulation need not be expected to treat each player symmetrically. Indeed should player 2 use the stipulation as a basis for deciding her expectation in swapping, she would find it unfavorable since she would find an expectation of $-8\frac{2k-1}{1+k}$ associated with the swap. That is, imagine both players have their unopened envelopes in hand. Player 1 opens his and calls out "I have \$16. Do you want to swap?" At that point, player 2 knows only that her envelope has \$8 or \$32. Recapitulating player 1's calculations, she would find that swapping carries an unfavorable expectation of $-8\frac{2k-1}{1+k}$.

It may still be hard to shake the impression that this computation of expectation ought not to favor player 1. To see that this is unproblematic, it is helpful to imagine what would happen in many repeated plays of the game. Repeated plays would result in a sequence of states something like

(\$2, \$1), (\$4, \$8), (\$16, \$8) , (\$1, \$2), (\$2, \$4), ...

So far all our frequency counts would bear symmetrically on player 1 and player 2. In assuming that player 1 has \$16, however, we restrict our consideration to that small subsequence of states containing just (\$16, \$8) and (\$16, \$32).[4] This injects an asymmetry. In this subsequence, exactly because of this asymmetry, if player 1 were to swap, player 1 would gain on average by the amount of the expectation. Correspondingly, if player 2 were to swap, she would lose on average by the same amount.

Here is one more way to see that these considerations do not contradict the symmetry of the game. Imagine that player 1 peeks in his envelope and sees \$16 and that player 2 peeks in her envelope and sees \$32, but neither knows the contents of the other player's envelope. Each will find a positive expectation in swapping. Player 1 computes an expectation of $8\frac{2k-1}{1+k}$; player 2 computes an expectation of $16\frac{2k-1}{1+k}$. But these assessments derive from differing assumptions. Player 1 is assuming that the two possible states of the world are (\$16, \$8) and (\$16, \$32). Player 2 assumes the two possible states of the world are (\$16, \$32) and (\$64, \$32). Naturally their differing assumptions properly lead to different assessments of the expectations. These differences would be supported by frequency considerations. In many repeated plays, consider those in which player 2 sees \$32 in her envelope. If player 2 were to swap in each game

© 1998 University of Southern California and Blackwell Publishers Ltd.

of that subsequence, then player 2 would gain on average $16\frac{2k-1}{1+k}$ with each play. There is no contradiction with player 1's considerations. The assumption that player 1 saw \$16 in his envelope would lead us to a different subsequence in which the average effect of swapping would be different.

4. *Defense of Claim (ii): Why Expectations Fail Player 1 if the Amount in his Envelope is Unknown*

According to claim (i), for each particular amount that player 1 may find in his envelope, player 1 has a positive expectation in swapping. It is easy to assume that this last result entails that player 1 has a positive expectation in swapping if the amount in his envelope is unknown and probabilistically distributed. But it does not – and therein lies the paradox. If the amount in player 1's envelope is unknown, then the expectation in swapping is given as a sum over the possible amounts that may be in the envelope. The set up of the game enjoins us to contrive these probabilities so that there is a symmetry between the players. But in so doing we destroy player 1's advantage and the very existence of an expectation for the swap, for the sum is non-convergent.

Assume that player 1 does not know how much is in his envelope. He knows only that the envelopes contain amounts in accord with the probability distribution (5). Then to compute his expectation in swapping, player 1 must consider the infinite matrix

State	(1,2)	(2,1)	...	$(2^n,2^{n+1})$	$(2^{n+1},2^n)$	...
Probability	p_0	p_0	...	p_n	p_n	...
e_1	1	2	...	2^n	2^{n+1}	...
e_2	2	1	...	2^{n+1}	2^n	...

Matrix 2: Payoffs when the contents of the envelopes is unknown

Player 1 now asks for his expectation in swapping. To find it he must sum the infinite series

$$E(e_2–e_1) = (2–1)P(1,2) + (1–2)P(2,1) + ... + (2^{n+1}–2^n)P(2^n,2^{n+1}) + (2^n–2^{n+1})P(2^{n+1},2^n) + ... \quad (7)$$

Series (7) is an alternating series. With such series there is a danger that

© 1998 University of Southern California and Blackwell Publishers Ltd.

the series will fail to converge. In such cases, different groupings of terms can lead to the illusion that the series sums to various different, definite values. The famous example is the series

$$1 - 1 + 1 - 1 + 1 - 1 + \ldots$$

One grouping leads to the illusion that it sums to 0:

$$(1 - 1) + (1 - 1) + (1 - 1) + \ldots = 0 + 0 + 0 + \ldots = 0$$

Another leads to the illusion that it sums to 1:

$$1 - 1 + 1 - 1 + 1 - 1 + \ldots = 1 - (1 - 1) - (1 - 1) - \ldots = 1 - 0 - 0 - \ldots = 1$$

It turns out that the series (7) is of exactly this type: alternating and non-convergent. By grouping its terms differently we can promote the illusion that (7) sums to different values. We can quickly see that the different arguments used in the paradox to arrive at conflicting conclusions amount to different ways of grouping the terms in (7).

4.1 THE SYMMETRY ARGUMENT: EXPECTATION IS ZERO

In this case we notice that $P(2^n,2^{n+1}) = P(2^{n+1},2^n) = p_n$ so that we can group the terms of (7) in such a way that the series seems to sum to 0

$$\begin{aligned} E(e_2–e_1) &= [(2–1) + (1–2)]\, p_0 + \ldots + [(2^{n+1}–2^n) + (2^n–2^{n+1})]\, p_n + \ldots \\ &= [0]\, p_1 + \ldots + [0]\, p_n + \ldots = 0 + \ldots + 0 + \ldots = 0 \end{aligned} \qquad (7a)$$

This grouping is inspired by the intuition that for each state that yields a gain if player 1 swaps, there is a corresponding and equally probable state in which player 1 loses that same amount. Therefore, the intuition runs, all the gains and losses balance to give an expectation of zero. Thus the grouping of terms in (7a) corresponds to expressing the expectation $E(e_2–e_1)$ as a sum of conditional expectations

$$\begin{aligned} E(e_2–e_1) = \ldots + [&E(e_2–e_1|e_2=2^n \text{ and } e_1 = 2^{n+1}) \\ &+ E(e_2–e_1|e_2 = 2^{n+1} \text{ and } e_1 = 2^n)]\, p_n + \ldots \end{aligned}$$

The symmetry assures us that the pairs of expectations in the square brackets sum to zero. Since such pairs exhaust all the possibilities, the series seems to sum to zero.

4.2 DOMINANCE BY PLAYER 1: A SWAP FAVORS PLAYER 1

In this case, we notice that we can group states into pairs sharing the same amount for envelope 1. That is, we form pairs of states $(2^n,2^{n–1})$ and

© 1998 University of Southern California and Blackwell Publishers Ltd.

$(2^n,2^{n+1})$ for each n. If we sum the terms corresponding to these pairs first, we rewrite (7) as the sum of infinitely many positive terms. The series seems to sum to a value greater than zero. In fact it diverges.

$$\begin{aligned}\text{Expectation} = {} & (2-1)P(1,2) + [(1-2)P(2,1) + (4-2)P(2,4)] + \\ & \ldots + [(2^{n-1}-2^n)P(2^n,2^{n-1}) + (2^{n+1}-2^n)P(2^n,2^{n+1})] + \ldots\end{aligned}$$

To see this behavior, we need only calculate the value of each term in square brackets and notice that each term is positive. The general term is:

$$\begin{aligned}&[(2^{n-1}-2^n)P(2^n,2^{n-1}) + (2^{n+1}-2^n)P(2^n,2^{n+1})] \\ &= (1/2)(1-k)k^{n-1}(2^{n-1}-2^n) + (1/2)(1-k)k^n(2^{n+1}-2^n) \\ &= (1/2)(1-k)(2k)^{n-1}(2k-1) \qquad (8)\end{aligned}$$

This term will be positive when $1/2<k<1$. So the expectation becomes a sum of positive terms:

$$\begin{aligned}\text{Expectation} = {} & (2-1)P(1,2) + (1/2)(1-k)(2k-1) + (1/2)(1-k)2k(2k-1) + \\ & \ldots + (1/2)(1-k)(2k)^{n-1}(2k-1) + \ldots \\ = {} & (2-1)P(1,2) + (1/2)(1-k)(2k-1)\,(1 + 2k + \ldots + (2k)^{n-1} + \ldots) \\ = {} & \infty \qquad (7b)\end{aligned}$$

for $1/2<k<1$.

This mode of summing series (7) corresponds directly to the argument that player 1 has a positive expectation in swapping for each amount that may be in his envelope. Therefore, the argument goes, he must have an expectation of gain if he swaps without concern for what is in his envelope. To see this notice that the square bracket term (8) can be rewritten in terms of player 1's expectation of gain in swapping (6b) conditioned on his having 2^n in his envelope

$$(1/2)(1-k)(2k)^{n-1}(2k-1) = E(e_2-e_1|e_1=2^n)\ P(e_1=2^n)$$

Combining, we see that the grouping in the sum of (7b) corresponds to

$$\begin{aligned}E(e_2-e_1) = {} & E(e_2-e_1|e_1 = 1)\ P(e_1 = 1) + E(e_2-e_1|e_1 = 2)\ P(e_1 = 2) + \\ & \ldots + E(e_2-e_1|e_1 = 2^n)\ P(e_1 = 2^n) + \ldots\end{aligned}$$

Thus $E(e_2-e_1)$ can be written as a weighted sum of the conditional expectations (1b). Since we know that each is greater than zero, it seems assured that $E(e_2-e_1)$ is greater than zero. But that conclusion requires that we accept (7b) as the correct way to sum the series (7). This is an illusion. While we can write (7) in this form, the sum has no more meaning

© 1998 University of Southern California and Blackwell Publishers Ltd.

and the expectation of swapping no more definite value than does the sum of the alternating series 1 – 1 + 1 – 1 + ...

4.3 DOMINANCE BY PLAYER 2: A SWAP FAVORS PLAYER 2

Because of the symmetry of the game, The complete analysis of section 4.2 can be repeated, but this time forming pairs of states $(2^{n-1},2^n)$ and $(2^{n+1},2^n)$. The resulting sum for (7) will diverge in the negative direction. Taking this sum corresponds to giving the same analysis as section 4.2 from the point of view of player 2.

4.4 COMPATIBILITY OF CLAIMS (I) AND (II)

The mathematical results of the previous sections demonstrate, I believe, the compatibility of claims (i) and (ii). However an aura of paradox may still remain. This aura derives from an appealing but fallacious inference which runs as follows. Player 1 holds his envelope in his hand and thinks: "For any definite amount that may be in this envelope, I have an expectation of gain if I swap. Therefore, no matter what is in the envelope, I have an expectation of gain if I swap." The fallacy is in the final step. If the amount in his envelope is unknown and assumed probabilistically distributed, his expectation of gain in swapping is a weighted sum of the conditional expectations. This sum fails to exist. There is no expectation.

These results on expectation translate directly into results on average gains in repeated plays. In repeated plays, there will arise some sequence of states. Consider the statement that player 1 has an expectation of gain of $\$8\frac{2k-1}{1+k}$ if he swaps when he sees his envelope contains \$16. This tells us something about the subsequence of states, each of which has \$16 as the contents of the first envelope. In this subsequence of states, if player 1 swaps in each case, then on average he will gain $\$8\frac{2k-1}{1+k}$ per play. Similar results follow for other specific amounts that may be in envelope 1. Each result points us to a different subsequence and assures us that, on average, player 1 gains if he swaps. If we consider in turn each of the possible amounts that may be in envelope 1, then our subsequences will exhaust all the states of the original sequence

Given these results, it is almost irresistible to conclude that a similar result holds for the entire sequence; that is, that player 1 gains on average if he swaps. But precisely that is *not* the case. The pathological probability distribution that leads to a non-convergence of expectation (7) also leads to a failure of the laws of large numbers. In many repeated plays, the average gain or loss that results from swapping will fail to settle down to a unique value. The existence of a favorable average in each subsequence fails to ensure a corresponding average in the sequence as a whole. Teased

© 1998 University of Southern California and Blackwell Publishers Ltd.

out in terms of averages of repeated plays, the paradox rests on the failure of the following inference. In repeated plays it is the case that[5]

> Player 1 gains if he swaps in those games in which he has $1 in his envelope.
> Player 1 gains on average if he swaps in those games in which he has $2 in his envelope.
> Player 1 gains on average if he swaps in those games in which he has $4 in his envelope.
> Player 1 gains on average if he swaps in those games in which he has $8 in his envelope.
> ...

But (astonishingly!) it does *not* follow from these results that

> Player 1 gains on average if he swaps in all games.

It is interesting to observe how a unique average fails to arise in this last case in the limit of many repeated plays. In this limit, the average gain remains probabilistically distributed and symmetrically so about zero because of the symmetry (2). But instead of condensing onto 0, as is the case with a law of large numbers, the probability mass spreads from 0 to ever larger positive and negative values of the average. More precisely, consider the interval of values for the average gain (–L,L), where L is any magnitude greater than zero, no matter how great. In Appendix A it is shown by elementary considerations that, in the limit of infinitely many plays, the probability is greater than 1/2 that the average gain lies outside this interval. One need not look far to see why there is no unique, limiting average. If one inspects matrix 2, one sees that for large n one moves to states of the world of very low probability. But these are states in which swapping produces either very large gains or very large losses for player 1. As play continues, these unlikely states will eventually be realized and they will so disrupt the accumulated totals as to prevent any stable average gain emerging, casting the average to positive or negative values well removed from zero.

The connection between the existence of a stable average and the convergence of the expectation is not so straightforward. As we shall see in section 6.4 below, the latter convergence is necessary for a stable average in the sense of the strong law of large numbers, but not for a stable average in the sense of the weak law of large numbers.

© 1998 University of Southern California and Blackwell Publishers Ltd.

5. *Strategies*

So far, we have computed expectations. How do these results bear on strategies that we may devise for the game? There is an expectation of gain if player 1 swaps with any fixed amount in his envelope. Might we use this to devise a strategy of play favorable to player 1? What dashes these hopes is the symmetry of the game. The distribution of money in the envelopes is effected symmetrically according to condition (2). This condition states that the probability distribution is invariant under exchange of envelopes. Therefore, in swapping unopened envelopes, player 1 does not alter the probability of each of the possible amounts in his envelope. We can infer from this a result for any probabilistic index of the desirability of swapping.[6] Because of the symmetry, any such index will show that swapping is equally favorable to player 1 and to player 2. If player 1 can gain an advantage only by disadvantaging player 2, it follows that player 1 can gain no advantage in swapping.

Player 1 can gain an advantage only by introducing circumstances that violate the symmetry of the game. A few examples of strategies favorable to player 1 illustrate this.

- Player 1 may contrive to play games only in which some specified amount – \$16, say – is placed in his envelope. All other games are nullified. If he always swaps, player 1 gains an average ofin repeated plays $8\frac{2k-1}{1+k}$

But this gain depends on an asymmetry. For player 2 sees all games nullified but those in which there is \$8 or \$32 in her envelope.

- Player 1 may contrive to play games only in which amounts from a specified set – say \$1, \$2, ... \$512 – are placed in his envelope. All other games are nullified. Player 1 gains on average from swapping always in repeated plays.

However the game has become asymmetric, for play is reduced to those cases in which player 2 may find \$1, \$2, ... , \$512, \$1024 in her envelope (and only the unfavorable subset of those with \$1024).

- Player 1 may decide to swap just in case he finds \$16 in his envelope; otherwise he stands pat. He will gain on average.

But he gains only because player 2 is agreeing to adopt an asymmetrical strategy. To accommodate player 1's strategy, she must swap just in case she has either \$8 or \$32 in her envelope and player 1 asks for a swap.

© 1998 University of Southern California and Blackwell Publishers Ltd.

This list may be continued – for example, the last strategy may be generalized to the case in which player 1 swaps just in case he has any amount from some predetermined set in his envelope. In each case, player 1 gains an advantage only at the disadvantage of player 2 and because she agrees to the introduction of an asymmetry into the play of the game. A strategy that preserves the symmetry of the game – for example, both player's swap on all plays – gives neither player any systematic advantage.

6. Exchange Paradoxes and their Properties

6.1 A GENERAL CHARACTERIZATION

It is not hard to see that the exchange paradox described here is one instance of a class of similar paradoxes. To develop a clearer notion of this class, I will sketch a general characterization of the circumstances under which these paradoxes arise. We assume that there are an infinity of states S_1, S_2, ... that may arise with probabilities p_1, p_2, ... We must choose between two actions a and b which result in the payoffs A_1, ... , B_1, ... shown in the matrix:

State	S_1	S_2	...	S_n	S_{n+1}	...
Probability	p_1	p_2	...	p_n	p_{n+1}	...
a	A_1	A_2	...	A_n	B_{n+1}	...
b	B_1	B_2	...	B_n	B_{n+1}	...

Matrix 3: Payoffs for a generalized exchange paradox

For convenience, we will assume that the payoffs are non-negative:

$$A_n \geq 0 \qquad B_n \geq 0 \tag{9}$$

for all n. This payoff structure will admit an exchange type paradox if two conditions are satisfied:

I. Symmetry: The structure treats actions a and b symmetrically in the sense that there is a bijection from the set of states to the set of states f: $S_n \rightarrow S_m$ such that, for all n and m

$$p_n = p_m, \qquad A_n = B_m \text{ and } B_n = A_m \tag{10}$$

II. Non-uniqueness: The expectation of exchanging action b for a is represented by a non-convergent series; that is

© 1998 University of Southern California and Blackwell Publishers Ltd.

$$E(b-a) = \sum_{n=1}^{\infty}(B_n - A_n)p_n \tag{11}$$

is non-convergent. Or, more generally, the expectation may be given by a series that is convergent but sums to different values if the terms are not just grouped differently, but reordered.

6.2 PARADOXICAL DECOMPOSITION

The symmetry condition I ensures that there is no systematic advantage in exchanging action b for a. However the non-uniqueness condition ensures that there will be paradoxical decompositions of the original decision problem that can be used to fallaciously suggest the desirability of one action over the other. That is, it will always be possible to partition the set of states $\{S_1, ...\}$ into infinitely many subsets

$\{S^m_1, S^m_2, ...\}$ with m = 1, 2, 3, ...

with a paradoxical property: if the decision problem is reduced to the case in which just members of any one set of the partition are considered, then there is an expectation of gain in exchanging b for a. But, since this holds for each of the sets of the partition, one is lured to conclude (fallaciously) that there is an expectation of gain in exchanging b for a when one recombines each subproblem to recover the original problem.

To see the paradoxical decomposition more explicitly, assume that the probabilities of states $S^m_1, S^m_2, ...$ are $p^m_1, p^m_2, ...$ and the corresponding payoffs for actions a and b are $A^m_1, A^m_2, ...$ and $B^m_1, B^m_2, ...$ respectively. The original decision problem, as represented in Payoff Matrix 3, is decomposed into infinitely many decision problems (for m = 1, 2, 3, ...) with payoff matrices

State	S^m_1	S^m_2	...
Probability	$\left(\frac{p^{m_1}}{p^{m_1}+p^{m_2}+\cdots}\right)$	$\left(\frac{p^{m_2}}{p^{m_1}+p^{m_2}+\cdots}\right)$	...
a	A^m_1	A^m_2	...
b	B^m_1	B^m_2	...

Matrix 4: Payoffs for a subdecision of the generalized exchange paradox

It is shown in the Appendix that it is always possible to effect this decomposition so that expectation within subproblem m is strictly positive

© 1998 University of Southern California and Blackwell Publishers Ltd.

$$E(b-a \mid \text{problem } m)\sum_n (B^m{}_n - A^m{}_n)\frac{p^m{}_n}{\sum_r p^m{}_r} > 0 \qquad (12)$$

Thus it is easy to presume that the sum (11) has a unique, strictly positive value given by the series

$$E(b-a) = \sum_m \left(E(b-a \mid \text{problem } m)\sum_r p^m{}_r \right)$$

But it does not since, by supposition II, E(b–a) has no unique value.

6.3 THE INEVITABILITY OF DIVERGENT EXPECTATIONS

In the original exchange paradox, the expectation $E(e_2–e_1)$ is non-convergent. However it also happens that the expectations for the individual contents of each envelope, $E(e_1)$ and $E(e_2)$ diverge.[7] This raises the complication of an entanglement of exchange paradoxes with the St Petersburg paradox. Might another variant of the exchange paradox avoid these infinite expectations? It is easy to show that this is not the case; that is, in any decision problem conforming to Matrix 3 and satisfying conditions I and II, the expectations E(a) and E(b) diverge. This is a generalization of the result proved by Nalebuff (1989, pp.178–79).

To see this, note that the series (11) cannot be absolutely convergent.[8] That is, since all the payoffs A_n and B_n are non-negative, the series $\sum_{n=1}^{\infty} B_n p_n + A_n p_n$ diverges. But the symmetry condition allows us to relabel indices so that this divergent sum can be decomposed into two equal parts

$$\sum_{n=1}^{\infty} B_n p_n + A_n p_n = \sum_{m=1}^{\infty} A_m p_m + \sum_{n=1}^{\infty} A_n p_n = 2\sum_{n=1}^{\infty} A_n p_n = 2E(a)$$

Hence it follows that the expectation of a, E(a) diverges and, by symmetrical reasoning, E(b) diverges as well.

6.4 NON-CONVERGENCE OF EXPECTATION AND NON-EXISTENCE OF UNIQUE LONG-TERM AVERAGES

In the case of the original paradox, condition II was satisfied and the average gain in many repeated swaps failed to converge towards a unique value. But in the general case, the lack of expectation required by condition II is not sufficient to ensure that the average gain in swapping action b for a fails to converge towards a unique value. The limiting behavior of this average varies with the different senses of “limiting”

© 1998 University of Southern California and Blackwell Publishers Ltd.

invoked in the laws of large numbers. In all cases, of course, from symmetry considerations, if the average does converge to a unique value, then the value must be zero. In the sense of the strong law of large numbers, applied to this game, we ask whether the quantity Av, the average gain from swapping in repeated plays, converges with probability one to zero, so that

$$P\left(\lim_{plays\to\infty} Av = 0\right) = 1$$

The non-existence of the expectation is sufficient to ensure that convergence in this stronger sense fails (Feller, 1971, p. 236, section VII.8). The weak law of large number requires only converge in probability, so that for any $\varepsilon>0$

$$\lim_{plays\to\infty} P(|Av|< \varepsilon) = 1$$

Non-existence of the expectation is not sufficient to ensure failure of convergence in this weaker sense. While convergence in this weaker sense does fail in the original exchange paradox, it turns out that one can construct variant forms that still staisfy conditions I and II but in which the average gain does converge to 0 in the weaker sense. An example is given in Appendix C.

7. Another Example of an Exchange Paradox: The Marble Game

The payoff structure in Matrix 3 and its attendant conditions I and II supply a template for the generation of exchange type paradoxes. The only real challenge is to embed the payoff structure in a plausible story so that the paradox has some seductive force. Here is an example that shows how the essential characteristics of the exchange paradox may be varied.

A player engages in a game in a casino in which both player and house arrive at scores by drawing colored marbles from an urn. The urn starts with a white and a black marble. The house draws one randomly, halting if the marble is white. If not, the black marble is returned and another black marble added as well. Drawing is continued in this way until the white marble is drawn. The house's score is the total number of drawings. The player's score is arrived at by the same scheme. The probability that house or player have scores n = 1, 2, 3, ... turns out to be given by

$$\text{P(house = n)} = \text{P(player = n)} = \frac{1}{n(n+1)} = \frac{1}{n} - \frac{1}{n+1} \qquad (13)$$

© 1998 University of Southern California and Blackwell Publishers Ltd.

The winner is the one with the higher score. The loser pays the winner a penalty equal to twice the square of the loser's score. Equal scores result in a draw with no payments.

Play proceeds as follows. The house establishes its score by drawing marbles, but the score is kept completely secret from the player. The player must then decide whether to play. The player reasons as follows.

> If the house has scored 1, then I cannot lose. For, if I score 1, it is a draw; if I score anything higher, I win. But if the house has scored higher than one – n say – I may win or lose. But a short calculation[9] shows that in every case my expectation is favorable, lying between 1 and n. No matter what the house's secret score, my expected gain is positive if I play. I will play!

This argument seems to show that the player has an advantage in playing.[10] But this result is paradoxical, for the actual play of the game is perfectly symmetrical. Both house and player generate their scores by symmetric mechanisms and there is a symmetry in the rules that determine who wins and the penalty paid.[11]

The formulation of the marble game paradox does not fit the template of section 5.1 precisely. For example, there are outcomes with positive and negative payoffs for the player, contrary to (9). The version described here is a simplified version of a marble game paradox that does fit the template but is a little less striking. In it, there is a treasury available that awards prizes. During play, the winner of each game draws the prize from this treasury in accord with the rule in place; the prize is not paid as a penalty by the loser. Condition (9) is now satisfied since there are no negative payoffs for losers. To compute a paradoxical expectation, we ask if the player would like to change places with the house. The above expectations arise if we then calculate the expected gain that would be realized by the house, if the house would forgo its score and swap with the player.

8. Conclusion

The force of the exchange paradox resides in the tension between the true claims (i) and (ii) of the introductory section 1. The resolution is that the force of (i) is illusory. It is powerless to overcome the symmetry of the game. Indeed, we have seen that (i) may enable the devising of a strategy favorable to player 1 only in so far as that strategy is allowed to introduce an element that violates the symmetry. The exchange paradox illustrates a failure of dominance of expectation. It supplies a case in which the expectations indicate that one action is more advantageous than another

© 1998 University of Southern California and Blackwell Publishers Ltd.

in each of infinitely many decision problems. Yet when we combine all the problems into one, the advantage vanishes.

Department of History and Philosophy of Science
University of Pittsburgh

NOTES

[1] I am grateful to John Earman, Brian Hill and Sandy Zabell for helpful discussion.

[2] The distribution (5) normalizes to unity since

$$\sum_{\text{all states}} p_n = \sum_{n=0}^{\infty} P(2^n, 2^{n+1}) + P(2^{n+1}, 2^n)$$
$$= \sum_{n=0}^{\infty} (1-k)k^n = 1 \text{ for } 1/2 < k < 1 \text{ as set in (4).}$$

Nalebuff (1989, p. 178) seems to be proposing the same distribution with $k = 1\sqrt{2}$, although on this assumption I have been unable to recover the constant in his distribution or some of the calculations that follow. Linzer's (1994, p. 418) distribution seems to correspond to (5) with k = 2/3.

[3] For n=0, the only possible state is $(2^0,2^1) = (1,2)$, so player 1 has an assurance of gaining 2–1=1 in swapping.

The frequencies of ($16, $8) and ($16, $32) would be roughly in the ratio of 1:k because of (6b).

[4] The "gains on average" is to be understood in the usual sense of a suitable law of large numbers, such a described in Manoukian (1986, pp. 77–78). For example, if player 1 has 2^n in his envelope, then the probability that the average return of many plays is within any arbitrary ε of the expectation approaches unity as the number of plays increases without limit.

[6] While the expectation may diverge, we may entertain other indices that are well behaved. For example, we may select to truncate the series (7) of terms that are of such low probability that we feel morally certain that they will not obtain. The result will be a moral expectation of the value of swapping. Or we may rescale the amounts gained in swapping with a non-linear utility function so contrived as to allow convergence.

[7] To see this notice that

$$E(e_1) = \sum_{n=0}^{\infty} (2^n + 2^{n+1}) p_n$$
$$= \sum_{n=0}^{\infty} P(2^n + 2^{n+1})(1/2)(1-k)k^n > (1-k)\sum_{n=0}^{\infty} (2k)^n = \infty$$

for 1/2 < k < 1 and similarly for $E(e_2)$.

[8] If (11) is non-convergent, its lack of absolute convergence is just the contrapositive of the usual form of the theorem that asserts that, if a series is absolutely convergent then it is also convergent. (Farrar, p. 59). If (11) is convergent but sums to different values under reordering of terms, then its lack of absolute convergence is just the converse of another standard theorem that asserts that absolutely convergent series do not alter their sums under reordering of terms (Farrar, p. 71).

[9] To see this, note that the expectation is

© 1998 University of Southern California and Blackwell Publishers Ltd.

$$\sum_{i=n+1}^{\infty}\frac{2n^2}{i(i+1)}-\sum_{i=1}^{n-1}\frac{2i^2}{i(i+1)}$$
$$=2n^2\left(\frac{1}{n+1}-\frac{1}{n+2}+\frac{1}{n+2}-\frac{1}{n+3}+\cdots\right)-2\left(\frac{1}{2}+\frac{2}{3}+\cdots+\frac{n-1}{n}\right)$$
$$=\frac{2n^2}{n+1}-2\left(\frac{1}{2}+\frac{2}{3}+\cdots+\frac{n-1}{n}\right)$$

For n = 1, the second term is zero and the expectation is 1. For n>1, the second term is non-zero but bounded as $\left(\frac{1}{2}+\frac{2}{3}+\cdots+\frac{n-1}{n}\right)\le n-\frac{3}{2}$ so that the expectation is greater than or equal to $\frac{2n^2}{n+1}-2\left(n-\frac{3}{2}\right)=\frac{2}{n+1}+1>1.$ The second term is also bounded by $\frac{n-1}{2}\le\left(\frac{1}{2}+\frac{2}{3}+\cdots+\frac{n-1}{n}\right)$ so that the expectation is less than or equal to $\frac{2n^2}{n+1}-\frac{2(n-1)}{2}=\frac{n^2+1}{n+1}<n.$ Combining, we have the expectation lying between 1 and n.

[10] Presumably the house can exploit such delusions to enable charging a hefty entrance fee!

[11] Another payoff rule gives a more striking result that I have suppressed in fear of unnecessary entanglement with the St Petersburg paradox. If the winner must pay the loser a penalty equal to the winner's score, then for each score n of the house, the player has an infinite expectation in playing. For n=1 the player has an assured win and his expectation is

$$\frac{0}{1.2}+\frac{2}{2.3}+\frac{3}{3.4}+\frac{4}{4.5}+\cdots=\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\cdots=\infty.$$

For n>1, a win is no longer assured, but his expectation remains infinite and is

$$\frac{-n}{1.2}+\frac{-n}{2.3}+\cdots+\frac{-n}{(n-1)n}$$
$$+\frac{n+1}{(n+1)(n+2)}+\frac{n+2}{(n+2)(n+3)}+\frac{n+3}{(n+3)(n+4)}+\cdots$$
$$=-n\left(\frac{1}{1}-\frac{1}{2}+\frac{1}{2}-\frac{1}{3}+\cdots+\frac{1}{n-1}-\frac{1}{n}\right)+\left(\frac{1}{n+2}+\frac{1}{n+3}+\frac{1}{n+4}+\cdots\right)$$
$$=-(n-1)+\left(\frac{1}{n+2}+\frac{1}{n+3}+\frac{1}{n+4}+\cdots\right)=\infty$$

[12] To see the limit in the last step of (A5), note that $\lim_{N\to\infty}(1-k^N)^{2^N/L}=0,$ since $\log(1-k^N)^{2^N/L}=(2^N/L)\log(1-k^N)=-(2^N/L)(k^N+k^{2N}/2+k^{3N}/3+\ldots)<-(2^N/L)k^N\to-\infty$ as $N\to\infty$.

[13] Inspection of (12) shows it differs only in a positive multiplicative constant used to normalize the probabilities.

© 1998 University of Southern California and Blackwell Publishers Ltd.

REFERENCES

Christensen, R. and Utts, J. (1992). "Bayesian Resolution of the 'Exchange Paradox'", *American Statistician* 46, pp. 274–76.

Farrar, W. L. (1938). *A Text-Book of Convergence* (Oxford: Clarendon).

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, 2nd edn (New York: Wiley).

Linzer, E. (1994). "The Two Envelope Paradox", *American Mathematical Monthly* 101, pp. 417–19.

Manoukian, E. B. (1986). *Modern Concepts and Theorems of Mathematical Statistics* (New York: Springer).

Nalebuff, B. (1989). "Puzzles: The Other Person's Envelope is Always Greener", *Journal of Economic Perspectives* 3, pp. 171–81;

—— "Puzzles", *Journal of Economic Perspectives* 2 (1988), pp. 150-51.

Rawling, P. (1994). "A Note on the Two Envelopes Problem", *Theory and Decision* 36, pp. 97–102.

Sobel, J. H. (1994). "Two Envelopes", *Theory and Decision* 36, pp. 69–96.

Zabell, S. L. (1988). *Bayesian Statistics 3: Proceedings of the Third Valencia Meeting, June 1–5, 1987* (Oxford: Clarendon), pp. 232–41.

APPENDIX A: AVERAGE GAIN FROM SWAPPING IN MANY PLAYS OF THE EXCHANGE PARADOX GAME

To determine how the probabilities of different average gains develop over time, define the random variable $X=e_2-e_1$ as player 1's gain in swapping envelopes. It follows from (5) that X is distributed as

$$P(X=2^n) = P(X=-2^n) = (1/2)(1-k)k^n \tag{A1}$$

for n = 0, 1, 2, ... The outcome of n swaps is represented by n independent variables X_1, ... , X_n, all with the same distribution as X. Let their sum be $S_n = X_1+...+X_n$. We now show that the probability mass for the average gain S_n/n spreads from 0 in the course of many plays; that is, for an L>0,

$$\lim_{n\to\infty} P\left(\frac{|S_n|}{n} > L\right) > \frac{1}{2} \tag{A2}$$

Informally (A2) is proved by showing that, in the limit of infinitely many plays, the probability exceeds 1/2 that there is a single big win or big loss – called M in the proof below – that by itself is sufficient to throw the average gain S_n/n outside the interval (–L,L).

To see (A2) this we note that $|S_n|/n > L$ whenever $|S_n| > nL$. We place a lower bound on the probability of the latter by noting that we can writing $S_n = M + R$, where M is the first member of the set of variables X_1, … X_n that has the greatest absolute value and R is the sum of the remainder.

We have

$$P(|S_n|/n>L) = P(|S_n|>nL) > (1/2)\, P(|M|>nL) \tag{A3}$$

The last inequality of (A3) follows from

© 1998 University of Southern California and Blackwell Publishers Ltd.

$$P(M>nL) = P(\ (M>nL \text{ and } R\geq 0) \text{ or } (M>nL \text{ and } R<0))$$
$$< P(M>nL \text{ and } R\geq 0) + P(M>nL \text{ and } R<0)$$

But from the symmetry of the distribution of the X_i we have

$$P(M>nL \text{ and } R\geq 0) = P(M>nL \text{ and } R<0)$$

so that

$$P(M>nL) < 2P(M>nL \text{ and } R\geq 0).$$

Similarly we have

$$P(M<-nL) < 2P(M<-nL \text{ and } R<0).$$

Now

$$\begin{aligned} P(|Sn| > nL) &= P(|M + R| > nL) > P(\ (M>nL \text{ and } R\geq 0) \text{ or } (M<-nL \text{ and } R<0)\ ) \\ &= P(M>nL \text{ and } R\geq 0) + P(M<-nL \text{ and } R<0) \\ &\geq (1/2)\ P(M>nL) + (1/2)\ P(M<-nL) = (1/2)\ P(|M|>nL) \end{aligned}$$

which is the last in equality of (A3).

We now need to determine the limiting behavior of $P(|M|>nL)$ for large n. As n becomes large, for any value of nL, we can find a positive integer value of N such that $2^N \geq nL$. Therefore

$$\lim_{n\to\infty} P(|M|>nL) \geq \lim_{n\to\infty} P(|M|>2^N) \tag{A4}$$

We have from (A1) that $P(|X|<2^N) = (1/2)(1-k)(k^{-N}+...+k^{N-1}+k^N) = 1 - k^N$ for N any positive integer so that

$$\begin{aligned} P(|M|>2^N) &= 1 - P(|X_1|<2^N)\ P(|X^2|<2^N) \ ...\ P(|X_n|<2^N) \\ &= 1 - (1 - k^N)^n = 1 - (1 - k^N)^{2^N/L} \\ &\to 1 \text{ as } N \to\ \infty \end{aligned} \tag{A5}$$

Combining (A3), (A4) and (A5) returns the principal result (A2).[12]

APPENDIX B: EXISTENCE OF PARADOXICAL DECOMPOSITION OF THE DECISION PROBLEM

There are many ways of partitioning the set of states of the general exchange paradox of section 6 so that the paradoxical condition (12) holds in all sets of the partition. Constructing one suffices to show that conditions I and II always allow such paradoxical decompositions. To see this, we note that the series (11) is either non-convergent or alters its sum under reordering of terms. It follows that the series must have terms with both positive and negative signs, where we take $(B_n - A_n)p_n$ for each n as an individual term. Call those with positive signs P_1, P_2, ... and those with negative signs N_1, N_2, ..., ignoring those terms with zero value. Because of the symmetry of the problem, it is possible to assign these labels to terms so that $P_n = -N_n$ for all n. So the series (11) can be represented as

© 1998 University of Southern California and Blackwell Publishers Ltd.

$$P_1 + P_2 + ... + N_1 + N_2 + ...$$

It is easy to see that this series (11) is not absolutely convergent. For, if (11) is non-convergent, its lack of absolute convergence is just the contra-positive of the usual form of the theorem that asserts that, if a series is absolutely convergent, then it is also convergent (Farrar, p. 59). And, if (11) is convergent but sums to different values under reordering of terms, then its lack of absolute convergence is just the converse of another standard theorem that asserts that absolutely convergent series do not alter their sums under reordering of terms (Farrar, p. 71). Thus the series

$$|P_1| + |P_2| + ... + |N_1| + |N_2| + ...$$

diverges to positive infinity. But since the Pn are positive and |Pn| = |Nn|, both for all n, it follows that the series

$$P_1 + P_2 + ...$$

diverges to positive infinity.

To form the partition all of whose sets satisfy (12), we proceed as follows. We form the first set with the state corresponding to N_1 and add in sufficiently many states corresponding to the terms P_1, P_2, ... , P_p, so that the sum $N_1 + P_1 + P_2 + ... + P_p$ is greater than zero. Since $P_1 + P_2 + ...$ diverges, only finitely many terms from $P_1 + P_2 + ...$ will be needed to ensure that the sum is positive. But since the expectation (12) for this first subproblem is proportional[13] to $N_1 + P_1 + P_2 + ... + P_p$, it follows that this expectation is positive. The second set of the partition is formed likewise. It consists of the state corresponding to N_2 and sufficient of those corresponding to P_{p+1}, P_{p+2}, ... to ensure that the expectation (12) is positive. Since finitely many terms of $P_1 + P_2 + ...$ are required for each set, the procedure can be continued to form the partition desired. States associated with zero valued terms may then be distributed arbitrarily among the sets of the partition since they will not affect satisfaction of (11).

APPENDIX C: AN EXCHANGE PARADOX IN WHICH THE AVERAGE GAIN IN SWAPPING CONVERGES TO ZERO

This appendix describes a variant form of the original exchange paradox in which both:

- the expected gain in swapping envelopes is represented by a non-convergent series;
- the average gain from many swaps converges to zero in the sense of the weak law of large numbers.

This variant exchange game is identical with the version described in section 2 above with the exception that the amounts in the envelopes – the possible states – are no longer

$$(2^n, 2^{n+1}) \text{ or } (2^{n+1}, 2^n) \text{ where } n = 0, 1, ...$$

but are now given as

$$(f_n k^{-n}, f_{n+1} k^{-(n+1)}) \text{ or } (f_{n+1} k^{-(n+1)}, f_n k^{-n}) \text{ where } n = 0, 1, ...$$

and the coefficients f_n are defined recursively as

© 1998 University of Southern California and Blackwell Publishers Ltd.

$$f_0 = f_1 = 1 \quad f_{n+1} = k(1/n + f_n) \quad \text{for } n>1 \tag{C1}$$

As before, we have for all n>0

$$P(f_n k^{-n}, f_{n+1} k^{-(n+1)}) = P(f_{n+1} k^{-(n+1)}, f_n k^{-n}) = (1/2)(1-k)k^n \tag{C2}$$

The recursive definition of f_n is so contrived that the gain in swapping envelope 2 for envelope 1 is given by

$$\begin{array}{ll} k^{-1} - 1 & \text{for state } (f_0 k^0,\ f_1 k^1) \\ -(k^{-1} - 1) & \text{for state } (f_1 k^1,\ f_0 k^0) \\ (1/n)k^{-n} & \text{for state } (f_n k^{-n}, f_{n+1} k^{-(n+1)}) \\ -(1/n)k^{-n} & \text{for state } (f_{n+1} k^{-(n+1)}, f_n k^{-n}) \text{ where } n>0 \end{array} \tag{C3}$$

If we define the random variable $X = e_2 - e_1$ as the gain in swapping envelope 2 for envelope 1 then it follows from (C2) and (C3) that

$$\begin{array}{l} P(X = k^{-1} - 1) = P(X = -(k^{-1} - 1)) = (1/2)(1-k) \\ P(X = (1/n)k^{-n}) = P(X = -(1/n)k^{-n}) = (1/2)(1-k)k^n \quad \text{where } n>0 \end{array} \tag{C4}$$

(A) EXPECTATION IS NON-CONVERGENT

The expectation of X is given as a sum of two terms

$$E(X) = E(X|X>0) + E(X|X<0)$$

But each of these terms is represented by a divergent series. The first, for example, is given as

$$\begin{aligned} E(X|X>0) &= (1/2)(1-k)\,[\,(k^{-1} - 1) + (1/1)k^{-1}k^1 + (1/2)\,k^{-2}k^2 + \ldots\,] \\ &= (1/2)(1-k)\,[\,(k^{-1} - 1) + 1/1 + 1/2 + 1/3 + \ldots\,] \to \infty \end{aligned}$$

Similarly

$$E(X|X<0) \to -\infty$$

It now follows immediately that E(X) is non-convergent; different orders of summing terms from the two subsequences E(X|X>0) and E(X|X<0) can be so contrived as to give different results.

(B) WEAK LAW OF LARGE NUMBERS HOLDS

The weak law of large numbers applied to this game asserts that the average A_N of gains from swaps in N plays converges to zero in probability; that is, for any $\varepsilon>0$

$$\lim_{N\to\infty} P(|A_N| < \varepsilon) = 1$$

The necessary and sufficient condition for this convergence to obtain is (Feller, 1971, p. 235)

$$\lim_{N\to\infty} [1 - F(t) + F(-t)]t = 0 \tag{C5}$$

© 1998 University of Southern California and Blackwell Publishers Ltd.

where F is the cumulative distribution $F(t) = P(X \prime\prime t)$. We have for $n>0$

$$F((1/n)k^{-n}) = (1/2) + (1/2)(1–k)(1 + k + ... + k^n) = 1 – (1/2)k^n$$
$$F(–(1/n)k^{-n}) = (1/2)(1–k)(k^n + k^{n+1} + k^{n+2} + ...) = (1/2)k^n$$

Setting $t=(1/n)k^{-n}$ and taking the limit $t=(1/n)k^{-n} \to \infty$ (for which $n \to \infty$), we have

$$(1/n)k^{-n} [1 – F((1/n)k^{-n}) + F(–(1/n)k^{-n})] = (1/n)k^{-n} [k^n] = 1/n \to 0$$

Condition (C5) is satisfied and the weak law of large numbers holds.

© 1998 University of Southern California and Blackwell Publishers Ltd.