Journal of Economic Perspectives— Volume 3, Number 1— Winter 1989—Pages 171–181

Puzzles
The Other Person's Envelope is Always Greener

Barry Nalebuff

The Puzzles feature in this issue is devoted to answering a puzzle that appeared in the Spring 1988 issue: "The Other Person's Envelope is Always Greener." The response to this puzzle was overwhelming. Below, I recap the problem and then discuss the myriad of proposed solutions. In the next column, we will return to the regular format of a few speed puzzles and one or two longer questions.

Contributions from readers are always valued; now they will be rewarded. We will begin providing *JEP* T-shirts to those contributing new puzzles and innovative answers. These T-shirts cannot be bought, only earned. Please send your answers, comments, and favorite puzzles to me directly: Barry Nalebuff, "Puzzles," Department of Economics, Princeton University, Princeton, N.J. 08544-1017.

The Problem

You have two envelopes. In one you place a hidden amount of money and give the envelope to Ali. Then you flip a hidden coin. If it comes up heads, you place twice the original amount of money in the second envelope. If it comes up tails, you put only half the original amount in the second envelope. You give this second envelope to Baba. So far, the contents of both envelopes are hidden, as is the outcome of the coin toss. Ali and Baba are allowed to look privately at the amount of money in their own envelopes. Then they are given an opportunity to trade envelopes if both agree.

Suppose, for the sake of argument, that Ali finds \$10.00 in her envelope. Ali reasons that Baba is equally likely to have \$5.00 or \$20.00. Trading envelopes gives

■ *Barry Nalebuff is an Assistant Professor of Economics, Princeton University, Princeton, New Jersey.*

her an expected gain of \$2.50 (or 25 percent). Acting in a risk-neutral manner, she would want to switch.

Now Baba looks inside his envelope. Whatever amount he finds (either \$5.00 or \$20.00), he too reasons that Ali is equally likely to have half or double his amount. The expectation is $0.5[0.5X + 2X] = 1.25X$, so he too expects a 25 percent gain from switching envelopes.

But this is paradoxical. The sum of the amount in both envelopes is whatever it is. Trading envelopes cannot make *both* participants better off. Yet, they both expect to make a 25 percent gain. Where did they go wrong?

In the "original" version of the problem, there is no coin toss. We are only told that one envelope contains twice as much money as the other, but not which is which. So before the envelopes are handed out, the two participants should be indifferent as to which they get. But once they open their envelopes, each appears eager to trade with the other. Just as above, both expect a 25 percent return from the trade. How can that be?

The History of the Problem

I want to thank Lawrence Waldman (Univ. of New Mexico) for writing in with an early description of the problem. In 1942, the French mathematician Maurice Kraitchik in his book *Mathematical Recreations* presents the "Paradox of the Neckties."

> Each of two persons claims to have the finer necktie. They call in a third person who must make a decision. The winner must give his necktie to the loser as consolation. Each of the contestants reasons as follows: "I know what my necktie is worth. I may lose it, but I may also win a better one, so the game is to my advantage." How can the game be of advantage to both? The question can be put in arithmetical form. Two people agree that they will compare the number of pennies in their purses, and that the one who has the greater number of pennies must give them all to the other. In the case of a tie, no money is transferred.

This first description of the paradox relies a little too heavily on there being an objective value of a necktie. Imagine that both players have a fifty-fifty chance of winning the other's tie. If the other tie is *truly* worth less to you (even when the referee says it's the nicer tie), then the expected value of this bet is negative. Thus, I prefer the purse exchange interpretation. Money is something we can objectively agree upon: more is better.

Kraitchik provides an answer to his own problem:

> From the point of view of the contestants the conditions of the game are symmetrical, so each has a probability of one-half of winning. In reality,

> however, the probability is not an objectively given fact, but depends upon one's knowledge of the circumstances. In the present case it is wise not to try to estimate the probability.
>
> We shall show that the game is no more advantageous for one than for the other. Let player A have *a* pennies in his pocket, and the other B have *b* pennies, and suppose that neither exceeds a certain fixed number *x*, say the total number of pennies minted to date. If all cases are equally probable, the possible gains of player A are the same as his possible losses, since the [payoff matrix is symmetric].

While I am happy to credit Kraitchik with the posing the question, I find his answer less than satisfactory. Martin Gardner, reviewing this puzzle in his book *Aha! Gotcha*, comments:

> Unfortunately, [Kraitchik's answer] does not tell us what is wrong with the reasoning of the two players. We have been unable to find a way to make this clear in any simple manner. Kraitchik is no help, and so far as we know, there is no other reference to the game.

I hope that by the end of this column, there will be no doubt as to the problem with the players' reasoning.

A second early statement of a closely related problem comes from the other side of the English Channel. Ian Jewitt (Bristol) describes a version of a puzzle appearing in J. E. Littlewood's (1953) *A Mathematician's Miscellany.*[1]

Littlewood puts the problem slightly differently. There is an infinite supply of cards; one is marked with a 0 on one side and 1 on the other, ten are marked with a 1 on one side and 2 on the other, ..., 10^n are marked with an n on one side and an $(n + 1)$, on the other, and so on. A card is drawn at random and held between two players A and B. The player seeing the higher number wins. It appears that both players believe their chance of winning is only one in eleven and thus both would prefer to have the other's side of the card.

To my mind, this version seems more artificial; in fact, Littlewood describes the distribution of cards as a "monstrous hypothesis." The reason is that, before we see the number on our card, we believe that the probability that the number will be less than N is zero *for any* N.[2] This statement of the paradox relies on infinity in a transparent

[1]It is reprinted in *Littlewood's Miscellany* (1986). In the 1986 edition, it appears as problem 4 on page 26, and Littlewood attributes it to the German physicist, Erwin Schrodinger.

[2]To see this, take any number, say 3. There are 1 + 10 + 100 + 1,000 cards with numbers 3 or below on one side. But there are 10,000 + 100,000 + 1,000,000 + ... cards imprinted with numbers above 3. The fraction of this infinite deck with numbers below any integer is zero.

way. There is no well-defined probability distribution which assigns probability zero to all numbers below N for all N!

Andrew Postlewaite (Univ. of Pennsylvania) relates a third variant of the problem, this one with the added fillip of a common knowledge twist. It is attributed to an early Martin Gardner column in *Scientific American*.

> Two people are given cards with consecutive positive integers. These cards are placed on their foreheads so that each person can see the card on the other's forehead but not the integer on his own. It is common knowledge that the cards are consecutive positive integers. One person is asked if he knows the integer on *his* card. If he says no, the other is asked. One can show by induction that eventually the person with the higher integer will be able to deduce the number on his forehead.[3]

Now we combine this common knowledge problem with the exchange paradox.

> Suppose we have the person who finally announces the number on his head collect that number of dollars from the other person. Now if I see the number 43, say, on my opponent's head, I know that I'll have either 42 or 44. If its 42, then I'll lose $43, while if its 44 then I'll win $44. Since it's equally likely that I'll have the higher of lower number, this is better than a fair gamble. Similarly, the other person expects to make money whatever he sees. Thus we have a pure transfer of money from one person to the other generating positive expected gains to both.

A appealing feature of this variant is that it works with additive rather than multiplicative payoffs. As we will see from the first response below, the multiplicative returns (twice or half) were a red herring that just provided additional complication.

And Now the Envelopes, Please: The Readers' Responses

I begin with a particularly innovative attempt at a solution. Lawrence Waldman provides an explanation that there is in fact no problem.

> I believe there is nothing wrong with the reasoning of either Ali or Baba. . . . Each is motivated to trade, expecting a long run gain. And the thing is, each

[3] Understanding the backwards induction argument is not essential to the puzzle. Still, it is easy to see for small numbers; i.e., if the numbers are 1 and 2, the person seeing 1 knows that this must be the higher number. If the numbers are 2 and 3, then when the person with 2 on his forehead does not know the answer it must follow that he has not seen 1, so that the other players now knows he has a 3. This is just a variant of the classic missionary problem, about which you will hear more in future *JEP* articles on common knowledge.

> will in fact show a long run gain of twenty–five percent. This is due to the fact that percentagewise, on any given trade, one stands to gain more (100 percent) than the other stands to lose (50 percent). So on each trade, the net percentage gain ... will be apportioned 25 percent to each. It is always the case that when two individuals trade items of unequal value, the percentage net gain is positive. Suppose A initially receives a, and B receives b, where $0 < a < b$. A calculates her proportional gain as $(b/a - 1)$. B calculates his as $(a/b - 1)$. The total gain is then $(a^2 + b^2 - 2ab) > 0$. Of course, in absolute terms, there will be no expected long run gain or loss.

Waldman's argument is that the paradox is one of perspective. Viewed in percentage terms there is room for mutual gain. Viewed in absolute terms, it is a zero sum game. The apparent contradiction is only a confusion of the two perspectives.

It is true that *before* either side opens his or her envelope they each expect a 25 percent gain but no absolute gain from trade. Since the bigger percentage gain comes on the smaller amount, a proportional gain on average for both parties should not be considered a paradox. Good point. But the puzzle is not about the sum of the expectations of the two parties: rather, why do they want to trade envelopes?

This approach cannot explain how both sides *simultaneously* believe they will gain 25 percent *after* looking at the amount in their envelopes. Once they know the amounts, each still expects a positive percentage gain and a positive absolute gain from trade: a contradiction.

Paradox Lost

The first solution of the problem is due to Sandy Zabell in his article "Loss and Gain: The Exchange Paradox." He argues that it is improper to assume that gains and losses are equally likely no matter what Baba sees in his envelope. Among my readers, Steven Lippman (UCLA) gets the credit for presenting this argument first. He explains the fallacy in Baba's reasoning is that Baba believes the amount he sees is *uninformative* with respect to the posterior probability his envelope contains the higher amount. That means that Baba believes that the probability his envelope contains the higher amount is $1/2$ regardless of what amount he sees in the envelope. This is true only if every value, from zero to infinity, is equally likely. But if an infinite number of possibilities are all equally likely, the chance of any one outcome must be zero. Then every outcome has a zero chance, and this is nonsense.

This argument can be made more formally. Let the prior probability that Ali's envelope contains an amount x be denoted by $f(x)$. When Baba observes an amount X, his posterior probability that Ali has $2X$ is (using Bayes rule)

$$\text{Prob}[\text{Ali} = 2X | \text{Baba} = X] = f(2X)/[f(X/2) + f(2X)].$$

In the statement of the problem, Baba is supposed to believe this probability is $1/2$ no matter what X he sees. Thus $f(X/2) = f(2X)$ for all $X > 0$. That requires $f(x)$ to be constant over $(0, \infty)$. There is no such thing as a uniform density over the real line. If the density is positive anywhere, the cumulative density would be infinite. Thus it must be zero, so the cumulative is always zero. No proper density function can lead to equal posterior probabilities for all observations.

Ian Jewitt puts it nicely. "As I see it, the monstrous hypothesis is the uniform density on the set of positive integers. Accept this and you accept anything." Two other readers recognized this problem and sent in equally elegant solutions: F. Trenery Dolbear Jr. (Brandeis) and Andrew Postlewaite (Univ. of Pennsylvania).

Paradox Buried

There are two other ways of looking at the problem that cast additional insight. One approach is based on the reality that the amount of money in the envelope must be bounded. As Kraitchik put it, if we are switching pennies, there is a limit to the total number of pennies ever minted and that places an upper bound on what we might expect to win.

Edward Norton, an MIT graduate student (and one of my undergraduate students at Princeton), details the effect of placing an upper bound.

> Suppose the arbitrator can never put more than A in the first envelope. Then the amount of money in Baba's envelope, b, must lie in the range $[0, 2A]$. Baba knows that when he finds b between $A/2$ and $2A$, the coin must have landed heads. Ali cannot have more than A, so that Ali's envelope must contain $a = b/2$. Therefore, Baba would never trade if he finds $b \geq A/2$, since his expected gain is negative. Ali then reasons that if she has between $A/4$ and A, she should not trade. Why? If Baba's envelope is larger, his b must be between $A/2$ and $2A$ and thus he will refuse to trade. The only time Baba would be willing to trade is when his envelope has between $A/8$ and $A/2$, in which case Ali loses money by trading. Similarly, once Baba recognizes that Ali won't trade when her envelope contains anything between $A/4$ and A, Baba should not want to trade when his envelope contains anything between $A/8$ and $A/2$. The reasoning continues inductively so that neither Ali nor Baba would ever want to trade.

This technique of solution based on the use of a bounded support was first proposed by Hal Varian. Several readers pointed out that with this restriction it could not be the case that Baba always wants to trade. But to get full credit, one has to take the inductive step and argue that hardly ever becomes truly never: the parties can never agree to trade.

The above argument shows the impossibility of trade when Ali and Baba agree that the maximum possible amount in either envelope is bounded by some high number.[4] Is it possible to recover the paradox by arguing that the upper bound should be considered infinite and thus nonexistent? More formally, is it possible to find some proper distribution on an unbounded support so that both Ali and Baba will always want to trade envelopes?

Professor Dolbear argues the answer is no. Here the explanation necessarily gets mathematical. "No probability density exists which guarantees that Baba will want to trade whatever value of x he discovers. Desire to trade given *any* x implies that Baba's prior probability distribution on the amount in Ali's envelope must drop over the range $0.5x$ to $2x$ by less than 50 percent." Otherwise, the expected return from trading is negative.[5] Once $f(2x) \geq .5f(x/2)$, we can place a lower bound on f: $f(x) \geq k\sqrt{x}$. The integral of this function diverges when x is unbounded and hence cannot represent a proper probability density.

This explanation is not a no with a capital N. Dolbear's argument requires that *any* amount of money might be found in the envelope. If instead, as seems reasonable, our prior beliefs are restricted to some integer amounts of money, potentially unbounded, the paradox may reappear.

Paradox Found?

There remains a way back to the paradox. Think back to the "original" statement of the problem where the envelopes are mixed before being handed out. Imagine that the amounts in the envelopes are *restricted* to the possibilities, 1, 2, 4, $8, \ldots, 2^n \ldots$ A person who sees 1 knows he has the lowest possible amount and would clearly want to switch. Is it possible that both Ali and Baba would always want to switch?

Yes! We construct a probability distribution so that after looking in the envelope Baba believes the odds are two-to-one that his is the higher amount. Let the prior

[4]They need not agree on the maximum bound. If one thinks the amount must be less than 1,000 and the other thinks it must be less than 3,000, both agree that it must be less than 1,000 and that is all that is needed for the argument.

[5]If $f(2x)$ is ever less than $.5f(x/2)$, the expected value of trading from x is negative so Baba will not offer to trade. Let

$$f(2x) < 0.5f(x/2) \Rightarrow R = f(2x)/f(x/2) < 1/2.$$

The expected value of trading is

$$[(x/2)f(x/2) + 2xf(2x)]/[f(x/2) + f(2x)] = (x/2) + (3x/2)R/[1 + R] < x$$

for $R < 1/2$.

probability that the minimum amount in the envelopes is 2^n be

$$\text{Prob}[\min(a, b) = 2^n] = K\sqrt{2}^{-n}, \text{with } k = \sqrt{2}/[2 - \sqrt{2}].$$

This is a well-defined probability density that sums up to 1. When Baba sees an amount b, his posterior probability that his envelope contains the larger amount is

$$\text{Prob}(b > a | b = 2^n) = \left[\sqrt{2}^{-n}\right] / \left[\sqrt{2}^{-n} + \sqrt{2}^{-(n-2)}\right] = 1/[1 + 2] = 1/3.$$

The odds are always two to one odds that he has the greater amount. When he sees Y, his expected payment from switching is

$$(2/3) \times Y/2 + (1/3) \times 2Y = Y.$$

Baba is indifferent and by the exact same reasoning so is Ali. (If the probability density falls slightly slower, then both sides will believe there is slightly less than a $1/3$ chance their envelope contains the greater amount and both *strictly* prefer to trade.)

It appears our original contradiction has returned. With probability one, both individuals want to trade. Although the paradox has returned, a new answer comes along side. Consider Baba's expected utility before opening the envelope. He is risk neutral, so that expected utility equals expected payoff which is

$$k\{1 \times 2^{-1/2} + 2 \times 2^{-1} + 4 \times 2^{-3/2} + \cdots \to \infty$$

Expected utility is unbounded. While the probability distribution is proper, the expected utility function is not. This is again nonsensical. If the expected utility of playing this game is unbounded, Ali and Baba should each be willing to pay an arbitrarily high amount for an arbitrarily small chance of finding out what lies inside their envelope. Here, we are reminded of the Saint Petersburg Paradox [see Samuelson (1977) for a defanging]. With the "monstrous hypothesis" of infinite expected utility, many things are possible.

So far, the argument is only by example. Now we show that if it is ever the case that the two sides always want to trade this implies expected utilities must be unbounded, whether or not the players are risk-neutral. Imagine that there exists some well defined probability distribution on the lesser amount, x_i. If trade is always to take place then there is an increasing series x_i such that whatever term Baba sees, he always prefers to take a gamble on Ali's envelope x_{i+1} or x_{i-1} rather than staying with x_i.

$$(1) \qquad \frac{p_{i-1}U(x_{i-1}) + p_{i+1}U(x_{i+1})}{p_{i-1} + p_{i+1}} > U(x_i)\ \forall i.$$

$$
\begin{aligned}
(2) \quad E[U] &= \sum_{i=0}^{\infty} p_i U(x_i) \\
&= \frac{1}{2}\left\{ \sum_{i=1}^{\infty} \left[p_{i-1}U(x_{i-1}) + p_{i+1}U(x_{i+1}) \right] + p_0 U(x_0) + p_1 U(x_1) \right\} \\
&\geq \frac{1}{2}\left\{ \sum_{i=1}^{\infty} \left[(p_{i-1} + p_{i+1}) U(x_i) \right] + p_0 U(x_0) + p_1 U(x_1) \right\} \\
&= E[U] + \frac{1}{2}\left\{ \sum_{i=1}^{\infty} \left[(p_{i-1} - p_i) U(x_i) + (p_{i+1} - p_i) U(x_i) \right] \right. \\
&\qquad \left. + p_1 U(x_1) - p_0 U(x_0) \right\} \\
&= E[U] + \frac{1}{2}\left\{ \sum_{i=1}^{\infty} \left[p_{i-1}(U(x_i) - U(x_{i-1})) \right. \right. \\
&\qquad \left. \left. - p_{i+1}(U(x_{i+1}) - U(x_i)) \right] \right\} \\
&= E[U] + \frac{1}{2}\left\{ \sum_{i=1}^{\infty} (p_{i-1} - p_i)(U(x_i) - U(x_{i-1})) \right. \\
&\qquad \left. + p_0(U(x_1) - U(x_0)) \right\} \\
&> E[U], \text{ a contradiction.}
\end{aligned}
$$

This constructive argument is a hard way of proving the no trade result for problems with common knowledge. As John Geanakoplos explains, if both parties want to trade no matter what they see *after* opening up their envelope, then they should be willing to trade *before* opening up their envelope. But that means they want to trade based on a common set of prior beliefs about what the envelopes contain. If the common prior leads to a well-defined expected utility, this is impossible, since the two parties must evaluate the envelopes in exactly the same way. Therefore, if Ali and Baba always want to trade, the common knowledge argument must not apply: the common prior beliefs must not lead to a well-defined expected utility.

The trick in using the common knowledge argument is to shift perspective from the ex post position once the envelopes have been opened to the ex ante position before information has been revealed. I recommend Sebenius and Geanakoplos (1983) for a simple exposition of one of these no-trade theorems.

Close But No Cigar

Finally, I want to comment on the most common response, especially since I don't find the argument entirely convincing. Ann van Ackere (London Business School), Frederick Fischer (Office of Management and Budget), Richard Green (Wisconsin Realtors Assoc.), and Arnold Kling (Federal Home Loan Mortgage Corporation) all wrote in with variants of the same approach. A representative argument goes as follows:

> Once the envelopes are handed out, the total amount of money in the two envelopes is fixed at some amount Z. Since one envelope contains twice as much money as the other, either you have $1/3$ of Z or $2/3$ of Z. Both cases are equally likely. Consequently, the expected value of what is in your envelope is $Z/2$. Since this is true for both players, they have no reason to trade.

The problem with this solution is that it requires that upon opening the envelope, the amount Baba sees is *uninformative* about the total Z. As we saw earlier, there is no proper prior distribution on Z such that the amount in Baba's envelope provides no information as to whether Ali's envelope contains the greater or lesser amount.

This explanation justifies why no trade should take place before the envelopes are opened, but, the paradox arises in the ex post perspective, once the two sides know how much they have. In particular, Baba may sometimes correctly expect a gain from trade for some values that he sees. Thus, this argument cannot explain what is wrong with Ali and Baba's proposed reasoning. In that way, it makes the original problem even more paradoxical.

A variant on the above approach was to apply the no trade result for a rational expectations equilibrium. Let Baba put himself in Ali's shoes. He should recognize that Ali has a 25 percent expected gain from trade. The reason is that a coin toss determines whether Baba's envelope contains half or double, the expectation of which is 1.25. Since Ali's calculation seems impeachable, Baba realizes that he will be on the other end of the trade and thus expects to lose 25 percent of the value. Once Baba puts himself in Ali's shoes, with rational expectations, he recognizes that Ali's expected gain is his expected loss.

The problem with this argument is that Ali's calculation is not so impeachable. Ali's expected 25 percent gain is conditional on Baba's *always* being willing to trade. When Baba is considering whether or not to trade in a particular instance, he cannot take as given Ali's expected payoff. Ali's expected payoff depends on Baba's choice of actions. For example, let the amount in the envelopes be an element of

$1, 2, 4, \ldots 2^n, \ldots$ When Baba sees 1, he knows that he has the least possible amount so that trading must be profitable.

Baba's ex ante calculation of Ali's expected gain is correct only if Baba is always willing to trade. This simply cannot be taken as given when Baba is deciding what to do. Moreover, as the above example illustrates, it does not apply to every ex post situation. But it does show that Baba cannot *always* be willing to trade. For then, Ali's ex ante calculation is correct and Baba should expect to lose money. Even here, caution is required. This argument moves us back to the ex ante perspective and there is no guarantee that expected utilities are well-defined.[6]

It is not that this second argument is wrong, but that it is incomplete. It doesn't tell us what is wrong with Ali and Baba's proposed reasoning. While it does show that trade cannot always take place, in fact, trade can never take place.

Perhaps a good paradox is never really eliminated, just dissected and defanged. I'm sure these explanations will provoke more correspondence from those who feel their solutions have been slighted. By viewing the paradox from all sorts of different perspectives, I hope everyone finds at least one of the arguments convincing. Martin Gardner's lament, at least and at last, has been well-addressed.

[6]In particular, we saw how this argument gets into trouble for the earlier example with risk-neutral players and where the probability of the lesser amount equals 2^n is $k\sqrt{2}^{-n}$; the ex ante expected utilities are unbounded.

References

Geanakoplos, John and James Sebenius, "Don't Bet on It," *Journal of the American Statistical Association*, 1983, 78, 424-426.

Samuelson, Paul Anthony, "St. Petersburg Paradoxes: Defanged, Dissected and Historically Described," *Journal of Economic Literature*, 1977, 15 (1), 24-55.