A SIMPLE SOLUTION TO THE TWO ENVELOPE PROBLEM

Ned MARKOSIAN

ABSTRACT: Various proposals have been made for solving The Two Envelope Problem. But even though the problem itself is easily stated and quite simple, the proposed solutions have not been. Some involve calculus, some involve considerations about infinite values, and some are complicated in other ways. Moreover, there is not yet any one solution that is widely accepted as correct. In addition to being notable for its simplicity and its lack of a generally agreed-upon solution, The Two Envelope Problem is also notable because it demonstrates that something that has been taken to be a fundamental principle of decision theory is false. The main purpose of this paper is to propose and defend a simple solution to The Two Envelope Problem. But I also want to make a start on the project of figuring out how the relevant fundamental principle of decision theory should be revised.

KEYWORDS: two envelope problem, decision theory, expected utility, decision under risk

1. The Problem

The Two Envelope Problem concerns a game – The Two Envelope Game – that goes like this. First, two seemingly indistinguishable envelopes, A and B, are presented to you. Each envelope contains some finite, non-zero amount of money, in dollars, and one envelope contains twice as much money as the other. Moreover, these facts about A and B are known to you. But you do not know how much money envelope A contains, or how much money envelope B contains, or even the total amount in the two envelopes. In this initial stage of the game, you are required to choose one of the envelopes, which then becomes your personal property. Then, in the second stage of the game, you are given the opportunity to trade your envelope for the other envelope. After the second stage of the game, you get to keep whichever envelope you are left with (together with its contents).

That's the whole game. It is a simple one, and a good one to play. Either way you are going to make some money.

Now, commonsense tells us that when you get to the second stage of The Two Envelope Game – where you get to decide whether to trade the envelope you have initially chosen for the other envelope – there is no special reason why you should trade. After all, given your evidence, you are just as likely to have chosen the envelope with more money as you are to have chosen the envelope with less money. And so, although it seems clear that you are rationally permitted to trade, in the second stage of The Two Envelope Game, it also seems clear that you are rationally permitted to stick with the envelope you initially chose.

But it turns out that there is an apparently sound argument, based on a fundamental and seemingly uncontroversial principle of decision theory, for the conclusion that it is rationally obligatory to trade in The Two Envelope Game.

The relevant principle of decision theory concerns the correct way to calculate expected utilities in what is known as a 'decision under risk' situation. A decision under risk situation is one in which the agent does not know with certainty what the outcomes of the relevant alternatives would be, but does know, for each outcome that might result from a given action, the *probability* of that outcome's resulting from that action, as well as the *value* of that outcome. Here is the principle.

> (FP) For any alternative, A, available to an agent, S, the *expected utility* of A can be calculated in the following way. First, determine all of the possible outcomes that might result from S's doing A. Second, for each such outcome, determine both the *probability* of that outcome (given that S does A) and the *value* of that outcome. Third, for each outcome, multiply the relevant probability times the relevant value. And finally, calculate the sum of all of these products. That sum = the expected utility of A.[1]

FP is meant to be one of the fundamental principles of decision theory (thus the abbreviation "FP"). It might at first appear to be complicated and somewhat arcane, but the idea that FP captures is actually very simple and intuitive. It would be hard to deny that this idea is on some level behind much of the reasoning that we (correctly) employ in everyday decision-making.

In any case, here is the argument from FP to the conclusion that it is rationally obligatory to trade envelopes when you get to the second stage of The Two Envelope Game. Let n = the amount of money in your envelope. Then, since there is a .5 probability that trading will mean trading up (thereby doubling your money), and a .5 probability that trading will mean trading down (thereby halving

---

[1] For principles relevantly like FP, see, for example, Brian Skyrms, *Choice and Chance: An Introduction to Inductive Logic,* 4th ed. (Belmont: Wadsworth, 2000), 128; and Michael D. Resnick, *Choices: An Introduction to Decision Theory* (Minneapolis: University of Minnesota Press, 1987), Ch. 3.

your money), FP entails that we can calculate the expected utility of trading as follows.[2]

expected utility of trading = [(.5 × 2n) + (.5 × .5n)] = 1.25n

Meanwhile, it is clear that if you stick, then you are guaranteed to end up with n (since that is the amount in your envelope). So FP entails that we can calculate the expected utility of sticking as follows.

expected utility of sticking = (1 × n) = n

Given these expected utilities (and assuming that it is always rationally obligatory to maximize expected utility), it looks like it is rationally obligatory to trade.

Since we know from commonsense that it is not rationally obligatory to trade in The Two Envelope Game, our problem is to say what, exactly, is wrong with this plausible argument.

Various proposals have been made for solving The Two Envelope Problem.[3] But, strangely enough, even though the problem itself is easily stated and quite simple, the proposed solutions have not been. Some involve calculus, some involve considerations about infinite values, and some are complicated in other ways. Moreover, there is not yet any one solution that is widely accepted as correct. My conviction is that a problem as simple as The Two Envelope Problem must have a simple solution. (And by 'a simple solution' I mean, roughly, one that does not involve any math more complicated than algebra.)

---

[2] Assuming, for the sake of simplicity, that the expected utility of any alternative = the expected dollar utility of that alternative.

[3] For more on the problem, see, for example, Barry Nalebuff, "The Other Person's Envelope Is Always Greener;" *Journal of Economic Perspectives* 3 (1989): 171-181; Frank Jackson, Peter Menzies, and Graham Oppy, "The Two Envelope 'Paradox,'" *Analysis* 54 (1994): 43-45; Paul Castel and Diderik Batens, "The Two-envelope Paradox: The Infinite Case," *Analysis* 54 (1994): 46-49; John Broome, "The Two-envelope Paradox," *Analysis* 55 (1995): 6-11; Timothy McGrew, David Shier, and Harry Silverstein, "The Two-envelope Paradox Resolved," *Analysis* 57 (1997): 28-33; Alexander D. Scott and Michael Scott, "What's in the Two Envelope Paradox?" *Analysis* 57 (1997): 34-41; Frank Arntzenius and David McCarthy, "The Two Envelope Paradox and Infinite Expectations," *Analysis* 57 (1997): 42-50; Michael Clark and Nicholas Shackel, "The Two-envelope Paradox," *Mind* 109 (2000): 415-442; Terry Horgan, "The Two-Envelope Paradox, Nonstandard Expected Utility, and the Intensionality of Probability," *Nous* 34 (2000): 578-603; David Chalmers, "The St. Petersburg Two-envelope Paradox," *Analysis* 62 (2002): 155-157; James Chase, "The Non-probabilistic Two Envelope Paradox," *Analysis* 62 (2002): 157-160; and Franz Dietrich and Christian List, "The Two-Envelope Pradox: An Axiomatic Approach," *Mind* 114 (2005): 239-248.

In addition to being notable for its simplicity and its lack of a generally agreed-upon solution, The Two Envelope Problem is also notable because it demonstrates that something that has been taken to be a fundamental principle of decision theory – namely, FP – is false. The main purpose of this paper is to propose and defend a simple solution to The Two Envelope Problem. But I also want to make a start on the project of figuring out how FP should be revised.

2. A Variety of Ways to Calculate Expected Utilities in The Two Envelope Game

Notice that we could just as easily reason about the expected utility of trading in The Two Envelope Game as follows. Let m = the amount of money in the other envelope. Then we get these calculations.

expected utility of trading = $(1 \times m) = m$

expected utility of sticking = $[(.5 \times 2m) + (.5 \times .5m)] = 1.25m$

Given this way of calculating expected utilities, it looks like it is rationally obligatory to stick.

Or we could reason as follows. Let x = the lesser of the two amounts in the envelopes. Then if you trade, there are two possibilities: either you will have x in your envelope and will double your money by trading, or you will have 2x in your envelope and will halve your money by trading. Whereas if you stick, then either you will be sticking with x or you will be sticking with 2x. Thus, we get these calculations.

expected utility of trading = $[(.5 \times 2x) + (.5 \times x)] = 1.5x$

expected utility of sticking = $[(.5 \times x) + (.5 \times 2x)] = 1.5x$

According to this way of calculating expected utilities it is neither rationally obligatory to trade nor rationally obligatory to stick.

There's more. We could also reason as follows. Let y = the greater of the two amounts in the envelopes. Then if you trade, there are two possibilities: either you will have y in your envelope and will halve your money by trading, or you will have .5y in your envelope and will double your money by trading. Whereas if you stick, then either you will be sticking with y or you will be sticking with .5y. We thus get these calculations.

expected utility of trading = $[(.5 \times .5y) + (.5 \times y)] = .75y$

expected utility of sticking = $[(.5 \times y) + (.5 \times .5y)] = .75y$

And according to these calculations, it is neither rationally obligatory to trade nor rationally obligatory to stick.

And finally, since there is one other significant quantity in the game that we could name, we could reason like this. Let z = the total amount in the two envelopes. Then one envelope contains 1/3z, while the other envelope contains 2/3z, so that if you trade, there is a .5 probability that you will trade up from 1/3z to 2/3z, and a .5 probability that you will trade down from 2/3z to 1/3z. Whereas if you stick, there is a .5 probability that you will be sticking with 1/3z and a .5 probability that you will be sticking with 2/3z. Thus, we get these calculations.

expected utility of trading = $[(.5 \times 1/3z) + (.5 \times 2/3z)] = .5z$

expected utility of sticking = $[(.5 \times 1/3z) + (.5 \times 2/3z)] = .5z$

According to this last way of calculating expected utilities, it is again neither rationally obligatory to trade nor rationally obligatory to stick.

We have seen that there are 5 important amounts of money involved in The Two Envelope Game that we can identify, and that, for each one, there is a corresponding way of calculating the expected utilities of trading and sticking.[4] These five ways can be summarized as follows.

Wn	n = the amount in your envelope. It is rationally obligatory to trade.
Wm	m = the amount in the other envelope. It is rationally obligatory to stick.
Wx	x = the lesser of the two amounts in the envelopes. It is neither rationally obligatory to trade nor rationally obligatory to stick.
Wy	y = the greater of the two amounts in the envelopes. It is neither rationally obligatory to trade nor rationally obligatory to stick.
Wz	z = the total amount in the two envelopes. It is neither rationally obligatory to trade nor rationally obligatory to stick.

[4] More accurately, there are five ways we can pick out an important amount of money involved in The Two Envelope Game – some pairs of which pick out the same amount of money (since, for example, the amount in your envelope may be the same amount as the lesser of the two amounts in the envelopes) – such that for each one, there is a corresponding way of calculating the expected utilities of trading and sticking.

So what should we do? Each of the five ways of calculating the relevant expected utilities seems to have a legitimate claim to being correct. Yet they yield results that are jointly inconsistent. Should we just observe that three of the ways yield the same result (that it is neither rationally obligatory to trade nor rationally obligatory to stick), while the other two 'cancel each other out,' and conclude that it is neither rationally obligatory to trade nor rationally obligatory to stick? Although that approach would give us the answer that we intuitively know to be correct, it doesn't really seem to be the right method for dealing with this kind of problem. Moreover, it will be recalled that our problem is not to show that, in The Two Envelope Game, it is neither rationally obligatory to trade nor rationally obligatory to stick, but, rather, to say what is wrong with the line of reasoning for the conclusion that it is rationally obligatory to trade in The Two Envelope Game.[5] In other words, what we need is some independent reason for claiming that Wn and Wm are wrong, while Wx-Wz are right. But is there such a reason?

3. Missing Information

I think there is. In particular, I think it can be shown that Wn and Wm are both incorrect ways of calculating the relevant expected utilities, and that Wx, Wy and Wz are all correct. The reason for this is that there is a certain crucial piece of information available to us about The Two Envelope Game that Wn and Wm (but not Wx-Wz) fail to take into account. Allow me to explain.

Consider a game that I will call *Double or Half*, or *DOH*. DOH is like the more traditional Double or Nothing, in that you start with some particular (finite, non-zero) sum of money, and you may choose either to play or not to play. If you decide to play, then some apparently indeterministic procedure with two possible outcomes that seem equally likely – such as a coin flip with a fair coin – is used to determine whether you win or lose. If you win, then your money is doubled, and if you lose, then your money is halved.

As any competent gambler will tell you, you should always jump at the chance to play DOH. For in DOH, the odds of winning are equal to the odds of losing, but you stand to win more than you stand to lose. I.e., if we let n = the amount of money you start with, then the following (familiar) calculations give us the expected utility of playing DOH and the expected utility of not playing DOH.

expected utility of playing DOH = $[(.5 \times 2n) + (.5 \times .5n)] = 1.25n$

expected utility of not playing DOH = $(1 \times n) = n$

[5] Presumably whatever flaw we identify in Wn will also appear in Wm.

In light of these expected utility calculations, we can see that the following principle is a good one.[6]

PlayDOH: It is always rationally obligatory to play DOH.

Now consider another game, which I will call *Thirds*. In Thirds, some particular amount of money is first selected, without your knowing what it is, and then that amount is divided into two smaller amounts, so that one smaller amount equals one-third of the total and the other smaller amount equals two-thirds of the total. Next comes the seemingly indeterministic part, which we will again suppose consists of the flipping of a fair coin. The coin flip in Thirds determines which of the two smaller amounts – one-third or two-thirds – is initially assigned to you. After that has been determined (but before you know whether you have been assigned one-third or two-thirds of the total), you get to decide whether or not you will trade your amount for the other amount. If you choose to trade, then you are choosing to play Thirds, and if you choose not to trade, then you are opting not to play Thirds.

If we let z = the total amount of money in the game, then here are the (again already familiar) expected utility calculations for playing Thirds.

expected utility of playing Thirds = $[(.5 \times 1/3z) + (.5 \times 2/3z)] = .5z$

expected utility of not playing Thirds = $[(.5 \times 1/3z) + (.5 \times 2/3z)] = .5z$

As these calculations show, it is neither rationally obligatory to play Thirds nor rationally obligatory not to play Thirds.

Now let's return to our Two Envelope Problem and the five ways of calculating expected utilities in The Two Envelope Game. Consider Wn, the original way of calculating expected utilities in The Two Envelope Game (and the way that was involved in the argument for the conclusion that it is rationally obligatory to trade). It should be clear that Wn is based on the assumption that when you trade in The Two Envelope Game, you are playing DOH. But this assumption is false. To trade in The Two Envelope Game is not to play DOH, because when you trade in The Two Envelope Game, it's not the case that, for some particular amount of money, a seemingly indeterministic procedure determines whether you will get double that amount or half that amount. Instead, what happens is this. First, an apparently indeterministic procedure determines whether you will initially receive one-third or two-thirds of some specific amount of money. After that, you are permitted to trade your amount for whichever one of the

[6] Again assuming that it is always rationally obligatory to maximize expected utility, and that expected utility = expected dollar utility.

smaller amounts you do not have. In other words, when you trade in The Two Envelope Game, you are playing Thirds rather than playing DOH. And that is why it is neither rationally obligatory to trade nor rationally obligatory to stick in The Two Envelope Game.

These observations about The Two Envelope Game, DOH, and Thirds allow us to see on an intuitive level not only what is right about Wx, Wy, and Wz (namely, that they are based on the observation that to trade in The Two Envelope Game is to play Thirds), but also what is wrong with Wn (namely, that it is based on the mistaken assumption that to trade in The Two Envelope Game is to play a game of DOH).[7]

Unfortunately, however, the above observations about The Two Envelope Game, DOH, and Thirds do not allow us to state in any precise way exactly why Wn and Wm are incorrect. For there is nothing to stop the proponent of Wn, for example, from giving the following speech.

> I agree with everything you've said about The Two Envelope Game, DOH, and Thirds. In particular, I admit that to trade in The Two Envelope Game is *not* to play DOH. But I never said it was, nor did anything in my calculations. All my calculations are committed to is this undisputed truth: As far as you know, there is a .5 probability that if you trade in The Two Envelope Game then you will be trading up (thereby doubling your money), and a .5 probability that if you trade in The Two Envelope Game then you will be trading down (thereby halving your money). Since that is all my calculations entail, and since, in particular, my calculations do not entail anything false (such as that you would be playing DOH if you were to trade in The Two Envelope Game), there is nothing wrong with my calculations.

A proponent of Wn who said this would be right in her claim that, as far as you know, there is a .5 probability that if you trade in The Two Envelope Game then you will be trading up (thereby doubling your money), and a .5 probability that if you trade in The Two Envelope Game then you will be trading down (thereby halving your money). But she would be wrong in claiming that her calculations do not entail anything false. For our thinking about DOH and Thirds has prepared the way for us to show, in a more precise fashion, exactly what is wrong with Wn and Wm.

First, let us note an obvious and uncontroversial fact about expected utility calculations. They are calculations of *expected* utilities. That is, they are calculations concerning how valuable the outcomes of various choices can be

[7] Similarly, the above observations about The Two Envelope Game, DOH, and Thirds allow us to see on an intuitive level what is wrong with Wm (namely, that it is based on the mistaken assumption that to stick in The Two Envelope Game is to play a game of DOH).

expected to be, *given the evidence available to us*. For this reason, it is possible for a way of calculating expected utilities to be incorrect simply because it fails to take into account some relevant piece of information that is available to us.

This fact about expected utility calculations will play an important role in my account of what is wrong with Wn and Wm. But in order to spell out the rest of that account, I will need to introduce another important fact about The Two Envelope Game. And in order to do that we will first need to introduce three technical terms. Let's say that you are playing a *trading game* iff you are playing a game in which, at some point, you have an opportunity to trade a certain amount of money that is already in your possession for some other amount of money. And let's say that you are playing a *fixed-total* trading game iff you are playing a trading game in which the total amount of money in the game (i.e., the amount initially in your possession + the amount that you can trade for) is fixed before it is determined how much money is initially in your possession. Finally, let's say that you are playing a *variable-total* trading game iff you are playing a trading game in which the total amount of money in the game is not fixed at the time when it is determined how much money is initially in your possession. Thus, for instance, Thirds is a fixed-total trading game (since the total amount of money in the game is already fixed by the time you are initially assigned your envelope), but DOH is a variable-total trading game (since the total amount of money in the game is not determined until after you are assigned your initial amount (and even then it is not determined unless you decide to trade)).

Now, here is a general fact about fixed-total trading games that can be easily proved. In any such game, the expected utility of trading = the expected utility of sticking = one half of the total amount in the game. And here is the proof. Suppose you are playing some fixed-total trading game. For any such game there are, before you decide whether to trade, three numbers, i, j, and k, such that the amount of money initially in your possession = i dollars, the amount that you are allowed to trade for = j dollars, and the total amount of money in the game (i.e., i dollars + j dollars) = k dollars. (This follows from the definition of a fixed-total trading game.) Then, since you have no reason to believe that you have either one of the two smaller amounts (namely, i and j) rather than the other, the expected utility of sticking is $(.5 \times i) + (.5 \times j)$. But that equals $.5 \times (i + j)$, and $(i + j) = k$. So the expected utility of sticking = .5k. And the same is true with regard to the expected utility of trading.

This important result about fixed-total trading games can be stated as follows.

> (IR) For any fixed-total trading game, G, the expected utility of trading in G = the expected utility of sticking in G = one half of the total amount in G.

It is important to note that IR is a purely general result. It applies to any fixed-total trading game, regardless of the specific amounts involved in the game, and regardless of the ratio between those amounts. Thus, no other information that is added to the fact that you are playing a fixed-total trading game can possibly affect this result. (With the exception, of course, of information that requires you to revise your estimation that you are playing a fixed-total game. But no such information is available to you in the case of The Two Envelope Game.) Indeed, even if you were allowed to open your envelope before deciding whether to trade in The Two Envelope Game, whatever you learned by doing so would not change the fact that you are playing a fixed-total game. It is for this very reason that opening your envelope in The Two Envelope Game does not change the fact that is is neither rationally obligatory to trade nor rationally obligatory to stick. And the simple reason for this is that learning the amount in your envelope does not change the fact that IR applies to your situation.

Let's return to the five ways of calculating expected utilities for trading and sticking in The Two Envelope Game. We can now see, given both the generality and the truth of IR, that the fact that the total amount of money in The Two Envelope Game is fixed before you get to decide whether to trade is of the utmost importance. And now we can also see what is wrong with Wn and Wm. Although Wx, Wy, and Wz all entail that the total amount in The Two Envelope Game is fixed (for Wx entails that the total amount = 3x, Wy entails that the total amount = 1.5y, and Wz entails that the total amount = z), neither Wn nor Wm does so (since all Wn entails about the total amount is that it is either 3n or 1.5n, and all Wm entails about the total amount is that it is either 3m or 1.5m).

The upshot is that Wn and Wm are defective, and give incorrect results, because they fail to incorporate a crucial fact about the situation in The Two Envelope Game that is available to us, namely, the simple fact that the total amount in The Two Envelope Game is fixed rather than variable.

4. Revising FP

Let's take stock. We have seen that FP – which has been taken to be one of the fundamental principles of decision theory – is in need of revision. For it can be used to yield false results in our situation involving The Two Envelope Game and, moreover, it yields inconsistent results in that situation. Meanwhile, I have argued that the problem with applying FP to our situation in The Two Envelope Game is that it is possible to calculate expected utilities in the way prescribed by FP while at the same time failing to take into account some information that is crucial to that situation.

These considerations suggest that we should simply add a clause to FP that explicitly says to take into account all relevant information, as in the following revised version of FP.

> (FP2) For any alternative, A, available to an agent, S, the *expected utility* of A can be calculated in the following way. First, determine all of the possible outcomes that might result from S's doing A. Second, for each such outcome, determine both the *probability* of that outcome (given that S does A) and the *value* of that outcome. *(But be sure to individuate these possible outcomes, probabilities, and values in such a way that no crucial information about S's situation is left out of the reckoning.)* Third, for each outcome, multiply the relevant probability times the relevant value. And finally, calculate the sum of all of these products. That sum = the expected utility of A.

It must be admitted that FP2, in virtue of that parenthetical remark in the second step, seems disturbingly inelegant, and lacks a certain precision that was an important feature of FP. It would be nice if we could come up with a variation on FP2 that gets around at least the second of these problems. But we have already seen that the correct solution to The Two Envelope Problem involves recognizing that we must revise FP, so that it says something about taking into account all relevant information. In light of this, I am inclined to think that the solution to The Two Envelope Problem is to replace FP with something along the lines of FP2.[8]

[8] Earlier versions of this paper were presented at the University of Massachusetts and The University of Rochester in 2005. I am grateful to members of both audiences, and to Ben Bradley, Stuart Brock, Kenny Easwaran, Adam Elga, Alan Hájek, Frances Howard-Snyder, Hud Hudson, Lee Markosian, Jonathan Schaffer, Ted Sider, Ryan Wasserman, and Catherine Wilson for helpful comments or discussion.