The two-envelope paradox and inference from an unknown

Jeff Speaks

June 24, 2002

The two-envelope paradox, or exchange paradox, is a reductio of a certain very intuitive sort of reasoning about situations in which one has only partial information. It shows that an agent can be in a situation in which she knows both that one of a series of propositions is true and that, were she to know which is true, she would be rational to ϕ, but still not be rational to ϕ. In other words, one cannot always infer from the claim that, were one to know some relevant piece of information, it would be rational for one to ϕ, to the conclusion that it is rational for one to ϕ *now*, on the basis of one's current information. That the two-envelope paradox is a paradox about such situations of partial information can be seen by distinguishing three versions of the paradox, which differ according to what information is possessed by the agent in the paradoxical situation.

1 Three versions of the paradox

The following facts are common to each of the three versions of the paradox. An agent begins with two envelopes in front of him. Each envelope contains some amount of money, and one contains twice the amount in the other. The agent knows these facts, but does not know which envelope contains the higher amount of money, or what the amounts are. In fact, the agent assigns equal probability to every assignment of values to the envelopes which meets the requirement that one envelope contains twice the amount in the other. Hence, for any two assignments of values to the envelopes which satisfies the requirement that one envelope contain twice the amount in the other, the agent assigns, conditional on the information that one of these two assignments of values to envelopes is the actual assignment, probability 0.5 to each.

A note is required at the outset about this core of information possessed by the agent in question. In what follows I shall go on to show that an agent possessed of this information can, by employing a certain kind of reasoning, arrive at patently false conclusions, and use this to show that there is a flaw in the sort of reasoning employed. But this assumes that this core of information is coherent; after all, it would be no surprise if an incoherent belief set led, by a perfectly valid chain of reasoning, to a false conclusion. And there is good reason to suppose that the belief set described in the preceding paragraph is incoherent: it requires that the agent assign equal probability to each of the infinitely many possible assignments of values to envelopes which meet the requirement that one contain twice the amount in the other, and there is no standard probability distribution on which each of a countable infinity of propositions is equiprobable, and the probabilities of those propositions sum to 1. The paradox can be reformulated without the assumption that the agent assigns equal probability to a countable infinity of propositions; but, because this reformulation makes it more difficult to grasp the intuitive force of the paradox, I shall delay discussion of this until §2 below. Ignoring for the moment the need for such a reformulation, the three versions of the paradox are as follows.

The open version of the paradox. In this version of the paradox, the agent knows not only the information outlined above, but also the amount of money in his envelope. Suppose that you're in the situation described above, and that you pick up one of the envelopes and find that it contains, say, \$20. You're then presented with the choice to switch envelopes, or keep the one you selected. Because you know that your envelope contains \$20, you now assign probability 1 to the proposition that the assignment of values to envelopes is either \$20/\$40 and \$10/\$20; after all, you assign probability 1 to the propositions that your envelope contains \$20, and that one envelope contains twice the amount in the other. Since, conditional on the information that one of two assignments of values to the envelopes has been made, you assign equal probability to each of these assignments, you now assign probability 0.5 to the proposition that the assignment is \$20/\$40, and probability 0.5 to the proposition that the assignment is \$10/\$20. Since you know that yours contains \$20, you assign probability 0.5 to the possibilities of the unopened envelope's containing \$40 and its containing \$10. Since the expected value of switching your envelope for the unopened envelope is then the average of these — \$25 — it seems that you should switch.

The hidden version of the paradox. You're in the same initial situation as before, but this time do not know the contents of either envelope. You select one of the envelopes, and are again presented, without looking in the envelope you've chosen, with the chance to switch. Should you? If you switched in the first case, the argument goes, you should switch here. After all, it doesn't matter what amount of money is inside the envelope; if there's an equal probability that the other contains double or half, then you stand to gain by switching. So you should switch. Suppose you do, and get the other unopened envelope. Without opening it, you're then presented with the choice to switch again. The problem is that it seems that you can go through exactly the same reasoning to arrive at the conclusion that you should switch again; and so on *ad infinitum*.

We can make this line of reasoning explicit as follows. There is some amount of money in the envelope which you've selected; call this amount, whatever it is, $\$X$. Then you should assign probability 1 to the disjunctive proposition that the assignment of values to envelopes is either $\$X/\$(2X)$ or $\$(0.5X)/\$X$. Since, conditional on the information that one of two assignments of values to the envelopes has been made, you assign equal probability to each of these assignments, you now assign probability 0.5 to the proposition that the assignment is $\$X/\$(2X)$, and probability 0.5 to the proposition that the assignment is $\$(0.5X)/\$X$. Since by stipulation yours contains $\$X$, you assign probability 0.5 to the possibilities of the unopened envelope's containing $\$(2X)$ and its containing $\$(0.5X)$. Expected utility of switching: $\$(1.25X)$. So it seems that you should switch. And, as above, it seems clear that if given a chance to switch again without looking in your new envelope, the same apparently compelling line of thought should lead you to switch again; and so on for subsequent offers to switch.

The reverse open version of the paradox. Your beliefs about the two envelopes with which you're presented are as above, and again you select one. But this time, the one you didn't select is opened, and you see that it contains \$20. Should you switch your unopened envelope for the \$20? Since, as above, you assign probability 1 to the proposition that one envelope contains twice the amount in the other, you should assign probability 1 to the disjunctive proposition that the assignment of values to envelopes is either \$20/\$40 and \$10/\$20. And since, also as above, you ought to, given the information that one of two assignments of values is correct, assign probability 0.5 to each, you ought to then assign probability 0.5 to the proposition that the assignment is \$20/\$40, and probability 0.5 to the proposition that the assignment is \$10/\$20. Since of course you know that the envelope not in your possession contains \$20, you assign probability 0.5 to the possibilities of the your unopened envelope's containing \$40 and its containing \$10. So the expected value of sticking with your unopened envelope is \$25, and the expected value of switching is, of course, \$20. You shouldn't switch, and not just because you're lazy; you would be rational in this situation to pay \$4 to keep the envelope you chose.

So we have three versions of the paradox, each accompanied by an intuitive argument concerning the rational course of action for an agent confronting that version of the paradox. Of these three, only the hidden version is clearly paradoxical; the hidden version of the paradox is the only version for which we know the intuitive argument given above to be unsound. We know this, not only because the conclusion that one should go on switching envelopes with a positive expectation of gain is absurd, but also because the case for switching in the hidden case is based on an extension of the case for switching in the open case. But, now that we have the reverse open version of the paradox on the table, it is clear that, by the same sort of extension of the reasoning which supported *not* switching in the reverse open case, we get an argument for not switching in the hidden version of the paradox.[1] So, since our arguments concerning the open and reverse open versions of the paradox, when extended to the hidden case, yield contradictory advice, we must reject either the initial arguments for switching and not switching, respectively, in the open and reverse open versions of the paradox, or we must find some problem with the extensions of those arguments to the hidden version of the paradox. I will defend the latter course.

2 The prior probability distribution

Now, as noted at the outset of the preceding section, these two ways of responding to the paradox do not exhaust the possibilities. One might claim that the source of the agent's irrational choice is not to be found in her reasoning about the situation — the way in which she updates her beliefs — but rather in her prior probability distribution.

As stated, each version of the paradox relies on the agent in question assigning equal probability to each of the countable infinity of assignments of values to envelopes which meet the requirement that one contain twice the amount in the other. But, the response goes, there is no standard probability distribution which assigns each of a countable infinity of propositions equal probability, and on which the probabilities of those propositions sum to 1; hence it is no surprise that we get paradoxical results from the belief set assigned to the agent: one of the agent's beliefs is incoherent, and this is the source of her paradoxical reasoning.[2]

It is important for me to head off this line of response to the paradox, since I want to argue that the paradox provides a reductio of a certain kind of

[1] The extension goes as follows. Call the amount of money in the envelope which you *don't have* $\$X$. Then you should assign probability 0.5 to the possibilities that the envelope in your possession contains $\$(2X)$ and that it contains $\$(0.5X)$. Expected utility of staying with what you have: $\$(1.25X)$ Expected utility of switching: $\$X$. So you should not switch.

[2] McGrew, Shier, & Silverstein [1997] object that such "mathematical" solutions are implausible, since the paradoxical reasoning surely involves some sort of "straightforward conceptual confusion" (29). But note that the proponent of the present solution need not regard ordinary subjects as making a complex mathematical mistake in their probability calculations; she need only regard the results of those calculations as flawed because of an assumption which is false for relatively complex mathematical reasons. There is nothing implausible about this sort of explanation of errors on the part of mathematically unsophisticated subjects.

reasoning. One cannot provide a reductio of a kind of reasoning by showing that an agent, whose beliefs are incoherent, would, by employing this reasoning, arrive at irrational courses of action; the irrationality of the agents actions in such cases can be explained by the incoherence of her beliefs as well as by the reasoning employed.

Fortunately, there is a way to reformulate the paradox, due to Broome [1995], which makes no use of the assumption that the agent assigns equal probability to each of an infinite number of propositions. To see how this works, recall what work in generating the paradox was done by the agent's prior probability distribution. The point was to have the agent's probability distribution be such that, on the basis of (i) finding a certain sum of money in one envelope, (ii) knowing that one envelope contains twice the amount in the other, and (iii) the probabilities assigned to different assignments of values to envelopes, the agent can update her beliefs by apparently cogent reasoning to arrive at the belief that the expected utility of taking the unopened envelope exceeds that of taking the open envelope. An assignment of equal probability to the propositions that the other envelope contains double or half achieves this end because then the choice to switch for the unopened envelope amounts to taking an even-odds double-or-half bet; and the expected utility of taking such a bet always exceeds the expected utility of refusing it. But there are other probability distributions which achieve the desired effect.

The prior probability distribution suggested by Broome begins with the stipulation that the only permissible assignments of values to envelopes be powers of 2. Then suppose that the agent's prior probability distribution assigns, for any non-negative integer n, probability $2^n/3^{n+1}$ to the proposition that the smaller of the two values in the envelopes is 2^n. So the agent assigns probability $\frac{1}{3}$ to the proposition that the smaller value is 1, probability $\frac{2}{9}$ to the proposition that the smaller value is 2, and so on. This probability function assigns probabilities to each of the permissible assignments of values to envelopes, and these probabilities sum to 1.

Now imagine having this prior probability distribution and opening your envelope to find some amount x of money, such that x is equal to 2^n, for some non-negative integer n. How should you calculate the expected value of switching? You know that the other envelope contains twice the amount in yours iff x is the smaller of the two values; and you know that the other envelope contains half the amount in yours iff $x/2$, or 2^{n-1}, is the smaller of the two values. Consider first the proposition that x is the smaller of the two values (and hence that the other envelope contains twice the amount in yours). Since you've learned that the lesser of the two values is either x or $x/2$, what we want is the probability you should assign to the proposition that the smaller of the two values (s) is x, conditional on the information that the lesser of the two values is either x or $x/2$, i.e. $Pr(s = x | s = x \vee s = x/2)$, or

$$\frac{Pr(s = x \,\&\, (s = x \vee s = x/2))}{Pr(s = x \vee s = x/2)}$$

which is equivalent to

$$\frac{Pr(s = x)}{Pr(s = x \vee s = x/2)}$$

Recalling that $x = 2^n$, then, using the agent's prior probability distribution which assigns, for any n, probability $2^n/3^{n+1}$ to the proposition that the smaller of the two values is 2^n, this is equivalent to

$$\frac{2^n/3^{n+1}}{2^n/3^{n+1} \;+\; 2^{n-1}/3^n} = \frac{2}{5}$$

Hence, finding any value x in your envelope, you should assign a probability of $\frac{2}{5}$ to the proposition that x is the smaller of the two values, and hence that the other envelope contains twice the amount in yours; you should assign probability $\frac{3}{5}$ to the proposition that $x/2$ is the smaller of the two values, and hence that the other envelope contains half the amount in yours. Finding any permissible value x in your envelope, should you switch? The expected utility of switching is

$$\frac{2}{5}(2x) + \frac{3}{5}(x/2) = \frac{11x}{10}$$

which is greater than x; so, as above, it seems that you should switch. This is all that is needed to generate the paradox; so we do not need to suppose that the agent assigns equal probability to each of an infinite number of propositions in order to generate the line of reasoning which seems to indicate that the agent should switch (and switch again, and again) in the hidden version of the paradox. So there is nothing incoherent in the set-up of the paradox; since the conclusion of the reasoning sketched above in the discussion of the hidden version of the paradox is unacceptable, there must be something wrong with this reasoning.

Might someone still object that the prior probability distribution ascribed to the agent in Broome's reformulation of the paradox is incoherent? As Broome notes, one can imagine situations in which the agent's probability distribution is not only coherent, but also intuitively the correct assignment. Suppose that we have a coin which comes up heads $\frac{2}{3}$ of the time, and tails $\frac{1}{3}$ of the time, and that our agent knows this. Then suppose that we inform the agent that dollar values are placed in envelopes according to the following system: we flip the coin, and if tails comes up, then \$1 is the lesser of the two values in the envelopes; if heads, we flip again. If tails comes up on the second flip, we let \$2 be the lesser of the two values; if heads, we flip again. We continue in this way until the coin comes up tails; if the coin comes up tails on the n^{th} toss, then the lesser of the two values in the envelopes will be 2^{n-1}. Since, given the way this coin works, the agent should assign to the proposition that the coin will come up tails on the n^{th} toss and not before a probability of $(\frac{2}{3})^{n-1}(\frac{1}{3}) = 2^{n-1}/3^n$, the agent should, in this case, have exactly the probability distribution described above. So we can design a version of the two-envelope paradox in which each of the relevant beliefs of the agent is rational and intuitively true, but where the agent seems to be trapped by the paradoxical reasoning nonetheless. This makes the task of solving the paradox all the more pressing.

3 Inference from an unknown

So we need to find some flaw in the extension of the argument for switching in the open version of the paradox to the argument for switching in the hidden version of the paradox. Recall the first two steps in this line of reasoning used to support switching in the hidden version of the paradox. They were:

> There is some amount of money in the envelope which you've selected; call this amount, whatever it is, $\$X$. Then you should assign probability 1 to the disjunctive proposition that the assignment of values to envelopes is either $\$X/\$(2X)$ or $\$(0.5X)/\$X$.

From there on, I claim, the reasoning employed really is parallel to that used in the open version; hence, if there is a mistake in the extension of the reasoning about the open and reverse open versions of the paradox to the hidden version, it must be found here.

The mistake comes in the transition from the first sentence to the second in the above passage. It does not follow from the fact that your envelope contains $\$X$ that you should assign probability 1 to this disjunctive proposition; the latter only follows from the supposition that your envelope contains $\$X$ *along with the supposition that you know that it contains this amount.* After all, if you didn't *know* how much the envelope contained, how could it be rational for you to assign probability 1 to just this disjunctive proposition, rather than any of infinitely many other disjunctive propositions whose disjuncts are two assignments of values to the envelopes which satisfy the requirement that one contains twice the amount in the other?

Compare: you select an envelope, but don't look to see how much is in it. In fact, the envelope contains \$20. So you are rational to assign probability 1 to the disjunctive proposition that the assignment of values to envelopes is either \$20/\$40 or \$10/\$20. This is clearly a mistake; you need to know that your envelope contains \$20 for your assignment of probability 1 to this proposition to be rational. But this is exactly analogous to the line of argument above. The need for this extra step is easy to miss in informal presentations of the paradox because it is easy to conflate the role of an agent describing an agent deciding between two envelopes, and the role of the agent who is deciding between the envelopes.

So the above passage, adding this extra step, should have read:

> There is some amount of money in the envelope which you've selected; call this amount, whatever it is, $\$X$. Suppose that you know that your envelope contains $\$X$. Then you should ...

In this sort of reasoning, one imagines a counterfactual situation in which one knows more than one actually knows, and uses facts about what it would be rational for one to do in this counterfactual situation in order to discern the rational course of action in the actual world. Specifically, one reasons that, whatever the amount of money in my envelope, were I to add the knowledge that that amount of money is in my envelope to my stock of information about this situation, it would be rational for me to switch (in that counterfactual situation). Hence it is rational for me to switch *now* (in my actual situation, without knowing what amount of money is in my envelope). What the two envelopes paradox shows is that this sort of reasoning about situations of partial information can lead one astray.

The inference characteristic of this sort of reasoning, which I shall call *inference from an unknown*, is the following:

$$\frac{A \text{ knows that } \exists p\ (p \text{ is true } \& \ (A \text{ learns } p \ \Box\!\!\rightarrow A \text{ is rational to } \phi))}{A \text{ is rational to } \phi}$$

I shall first show the role played by inference from an unknown in generating the paradox, and then show that the hidden version of the paradox provides a reductio of this sort of inference.

In reasoning about the hidden version of the paradox, we begin with the fact that an agent A knows that there is some value x of money in his envelope. Letting 'E' abbreviate 'is in A's envelope', this part of A's knowledge can be expressed as

$\exists x Ex$

The agent further knows that, for any value of money in his envelope, were she to know that value, she would be rational to switch envelopes;[3] that is, she knows that

$$\forall x\ (Ex \rightarrow (A \text{ learns that } Ex \ \Box\!\!\rightarrow A \text{ is rational to switch envelopes}))$$

We may assume that from these two propositions the agent is in a position to know the following logical consequence of them:

$$\exists x\ (Ex\ \&\ (A \text{ learns that } Ex \ \Box\!\!\rightarrow A \text{ is rational to switch envelopes}))$$

How are we to get from this bit of knowledge to the conclusion that A would be rational to switch envelopes in the hidden version of the paradox? By employing an instance of inference from an unknown, in this case

$$\frac{\exists x\ (Ex\ \&\ (A \text{ learns that } Ex \ \Box\!\!\rightarrow A \text{ is rational to switch envelopes}))}{A \text{ is rational to switch envelopes.}}$$

But if we reject the validity of inference from an unknown, there is no valid route from the agent's stock of knowledge to the conclusion that he should switch envelopes; hence there is no general argument for switching in the hidden version of the paradox.[4]

[3] She knows this by knowing the reasoning which supports switching in the open version of the paradox, and seeing that this reasoning generalizes to any value which she might find in her envelope.

[4] We're now in a position to see the importance of the extra premise added to the informal argument for switching in the hidden version of the paradox above. Were we to allow that, for any value X, if there are \$$X$ in the agent's envelope then the agent is rational to assign probability 1 to the proposition that the assignment of values to envelopes is either $X/2X$ or $0.5X/X$, — and that the agent knows this — then we could formalize the agent's reasoning as follows:

$$\frac{\begin{array}{l}\forall x(Ex \rightarrow A \text{ is rational to switch envelopes})\\ \exists x\ Ex\end{array}}{A \text{ is rational to switch}}$$

So far this is just to clarify an assumption needed to generate the paradox; in order to resolve the paradox we also need to show that this assumption — that inference from an unknown is valid — is false. The two envelope paradox provides a reductio of the validity of inference from an unknown because there are two relevant unknowns: not only the amount of money in the envelope which the agent has chosen, but also the amount in the envelope which she has not chosen. The key point is that the relation of the latter unknown to the decision whether or not to switch is exactly the reverse of the former. In the argument for switching in the hidden version of the paradox discussed above, one infers facts about what one ought to do in the actual world from facts about what one ought to do in a certain class of possible worlds; it is unsurprising that one ought to switch in these possible worlds, since they are just possible worlds in which one is confronting the open version of the paradox.[5] But in reasoning about the hidden version of the paradox we might as well have considered the class of possible worlds in which one knows the contents of the envelope one has *not* chosen. Since in these worlds one is confronting the reverse open version of the paradox — and, as we saw at the outset, one should not switch envelopes when facing this version of the paradox — one should not switch envelopes in these worlds. But, by using inference from an unknown, we can infer from this that one ought not to switch envelopes in the actual world, where one does not know the contents of either envelope. So the hidden version of the paradox provides a situation in which, using inference from an unknown, we can derive contradictory results: depending upon which unknown fact one imagines oneself knowing, one can derive either the conclusion that one ought to switch envelopes

which is valid and does not require inference from an unknown. The response to this argument was to deny the first premise; to make the first premise true we must reformulate it as

$$\forall x\ (Ex \rightarrow (A \text{ learns that } Ex \;\Box\!\!\rightarrow A \text{ is rational to switch envelopes}))$$

But to get from this claim to the conclusion that A ought to switch envelopes, we do need to employ inference from an unknown.

[5]Recall that the open version of the paradox is distinguished from the hidden version only by the fact that in the former one knows the contents of the envelope one has chosen; the possible worlds under consideration are just those worlds in which one adds to one's stock of knowledge the knowledge that one's envelope contains a certain amount of money.

in the actual world, or the conclusion that one ought not to switch envelopes.[6]

4 Can the paradox be reformulated without inference from an unknown?

So the solution suggested for the two-envelope paradox is to reject a form of reasoning needed to generate the paradox. Even if this solves the version of the paradox discussed above, one might still wonder whether the paradox can be generated without making use of inference from an unknown.

One attempt to do so would be to begin with the general knowledge possessed by the agent, and see what facts about the relative values in the envelopes the agent is in a position to deduce. Using 'Pr' as above to stand for the probability function of the agent in the situation, 'A' as a monadic predicate of amounts of money which holds of a value iff that value is the amount of money in the envelope initially selected by the agent, and 'B' as a similar predicate which holds of a value iff that value is the amount of money in the envelope *not* initially selected by the agent, we can represent the agent's knowledge as follows. First, since the agent knows that there is some amount of money in his envelope, he knows that

$$\exists x A(x)$$

Given the agent's prior probability distribution, discussed in §2 above, the agent should also assign the following probabilities to possible values of the envelopes:

$$\forall x\, [Pr(B(2x)|A(x)) = 0.4 \;\&\; Pr(B(0.5x)|A(x)) = 0.6]$$

Combining these two bits of information, the agent may arrive at

[6]There is an interesting parallel between inference from an unknown and the sort of reasoning employed in the prisoner's dilemma. There one does not know what the other prisoner has done, and this is a fact potentially relevant to the decision whether or not to inform on your fellow prisoner. One arrives at the decision to inform on him by reasoning that, whatever he has done, one would be better off informing; and he follows the same reasoning. I don't know whether this sheds any light on the prisoner's dilemma.

$$\exists x\,[A(x)\ \&\ Pr(B(2x)|A(x)) = 0.4\ \&\ Pr(B(0.5x)|A(x)) = 0.6]$$

which, conditionalizing, yields conclusion **C**:

C $\quad \exists x\,[A(x)\ \&\ Pr(B(2x)) = 0.4\ \&\ Pr(B(0.5x)) = 0.6]$

This seems as though it should be sufficient to convince the agent to switch envelopes; and, to arrive at it, we have not had to use inference from an unknown. So this seems to be a version of the paradox which avoids the response suggested above.

This version of the paradox, though, is really just a case of an agent failing to use all of the information at his disposal and, for this reason, arriving at a conclusion which is irrational given his initial information. Recall that the agent in question knows not only that there is some amount of money in his envelope, but also that there is some amount of money in the envelope that he did not choose; so he knows that

$$\exists x B(x)$$

Since his information about the amounts of money in the two envelopes is symmetrical, he also knows that

$$\forall x\,[Pr(A(2x)|B(x)) = 0.4\ \&\ Pr(A(0.5x)|B(x)) = 0.6]$$

But then he is in a position to derive a conclusion which is the converse of the above, namely conclusion **C***:

C* $\quad \exists x\,[B(x)\ \&\ Pr(A(2x)) = 0.4\ \&\ Pr(A(0.5x)) = 0.6]$

Were the agent possessed of initial information which allowed him to derive **C** but not **C***, then the agent would indeed be rational to switch envelopes; but, because he is in a position to derive both, and because the conjunction of **C** and

C* provides no reason to switch envelopes — the information this conjunction provides about the relative values of the envelopes privileges neither envelope — he has no reason to switch envelopes.

The basic point is a simple one. There is nothing paradoxical about a case in which an agent begins with several different pieces of knowledge, and, making faultless inferences but excluding some relevant portions of that knowledge, arrives at a decision to act which is irrational, given the totality of his initial knowledge. Consider a case in which an agent is rational in believing that the Reds win 75% of their games against the Cubs, and is rational in believing that they win a mere 25% when Wood is pitching. One day the Reds and playing the Cubs, Wood is pitching, and our agent knows these facts. He's then offered a $20 even odds bet that the Reds will win. He reasons as follows: "The Reds win three times as often as they lose against the Cubs; and I stand to lose no more than I stand to gain. So my expected utility of betting my $20 is $35; and my expected utility of not betting is only $20. So I should take the bet." The agent has made a mistake, but there was no flaw in the reasoning which led him to believe that his expected utility of betting is higher than that of not betting. But of course there's no paradox here; the agent's mistake is caused by his failure to take into account all the information he possesses which is relevant to his decision. The argument for switching in the hidden version of the paradox which does not employ inference from an unknown is analogous: just as in the case of the bet on baseball the agent ignores the relevant information that Wood is pitching, so in the case of the two-envelope paradox the agent ignores the information that he can deduce from the knowledge that there is some amount of money in the envelope which he has not chosen.

5 A residual paradox

So arguments for switching in the hidden version of the paradox make one of two mistakes: either they rely on the faulty principle of inference from an unknown, or they are simply cases in which the agent fails to use all of the information at his disposal. But even if this is correct, one might reply that this is only a partial solution to the two-envelope paradox: this response, after all, applies only to the hidden version of the paradox. But what of the open and reverse open versions of the paradox?

There are really two questions to answer here. First, are the intuitive arguments given above in §1 — that one ought to switch envelopes in the open version of the paradox, and ought not to switch envelopes in the reverse open version — correct? And second, if they are, how *could* they be? How, that is, could finding out how much money is in one of the envelopes possibly be information relevant to the decision whether or not to switch?

It seems to me that the intuitive argument for switching in the open version of the paradox is correct.[7] Given the agent's prior probability distribution and the information that the envelope he has chosen contains \$20, the sorts of calculations run through in §2 show that the agent should assign probability 0.4 to the proposition that the other envelope contains \$40, and probability 0.6 to the proposition that it contains \$10. Since this gives us an expected utility of switching of \$22, the agent should switch.

The real problem is in explaining how this conclusion can be correct. Let $t-1$ be a time after the agent has selected one of the two envelopes, but before he has opened it, and let t be a time after he has opened his envelope and found that it contains \$20. We know that at $t-1$ the agent is not rational to switch envelopes, but that at t he is rational to switch. The problem is that at $t-1$ the agent already knows that, no matter what value he finds in his envelope, he will be rational to switch at t; given this, how can the information that his envelope contains \$20 have any effect on what it is rational for him to do?

This is indeed puzzling. But I take it to be the lesson of the two-envelope paradox that such cases are possible: cases in which there is a gap between knowing that it would be rational to ϕ given knowledge that a particular disjunct of a true disjunction is true, and knowing that it is rational to ϕ.

[7]The argument for not switching in the reverse open version is exactly analogous; the remarks in this section should be taken to apply to each.

Bibliography

Broome, John [1995]. "The Two-Envelope Paradox." In *Analysis* 55:1, 6–11.

McGrew, Timothy, David Shier, & Harry S. Silverstein [1997]. "The Two-Envelope Paradox Resolved." In *Analysis* 57:1, 28–33.