Logique & Analyse **200** (2007), 359–365

A THREE-STEP SOLUTION TO THE TWO-ENVELOPE PARADOX

IGOR DOUVEN

Abstract
The two-envelope paradox is a paradox in game theory. This paper offers a solution to it in three steps. The first step is to recognize that the paradox is dependent on how one represents the game-theoretic situation that gives rise to it. In the second step, we note that the representation used in the paradox violates a familiar symmetry requirement. And in the third, we derive the solution from the symmetry requirement plus the given that we are dealing with a zero-sum game.

You and a colleague are shown two envelopes, and you are informed that one of them contains twice the amount of money that is in the other. You are then each given one of the envelopes, and you are given the opportunity to swap. What should you do? The following argument counsels swapping:

> Let the amount of money in your envelope be $\$a$. It is as likely as not that the other envelope contains twice the amount that is in yours as that it contains only half that amount. In the first case, by swapping you stand to gain $\$a$. In the second, by swapping you will lose $.5\$a$. So the expected gain of swapping envelopes is $.5 \times \$a + .5 \times -.5\$a = .25\$a$. Patently, the expected gain of keeping your envelope is $\$0$. You thus should take the opportunity to swap.

But surely something must be wrong here, for the same reasoning would have applied had you been given the other envelope, and in that case swapping would have resulted in your holding the very envelope you now have in your hands!

This is a version of the so-called two-envelope paradox. It is important to distinguish this version from a very similar one in which you are supposed to open your envelope before deciding on swapping. There is a consensus that the solution to the latter must involve making a clear distinction between the prior probability that your colleague’s envelope contains twice the amount in yours and the posterior probability, conditionalized on the information obtained by looking what is in your envelope.[1] Quite evidently, this approach cannot be applied to the paradox depicted above. In fact, what the right approach to the “no looking” version is, is still very much an open question. In the following, I focus on this version and propose what I believe to be a very simple solution to it. The solution is in three steps. The first step is to recognize that the paradox is representation-dependent in the sense that not all ways of formally representing the above game-theoretic situation or, equivalently, framing the relevant expected gain calculation, lead to paradox. In the second step, I note that the representation used in the paradox violates a familiar symmetry requirement. And in the third, I derive from the given that we are dealing with a zero-sum game the conclusion that a representation is non-paradoxical if it obeys the symmetry requirement.

1. Contrast the above, paradoxical argument with the following:

> Instead of calculating your expected gain in terms of the amount of money in your envelope, calculate it in terms of the total amount of money involved. Call this amount $\$b$. Then either you have $1/3\$b$ in your envelope and your colleague has $2/3\$b$ in hers, or you have $2/3\$b$ in your envelope and she has $1/3\$b$ in hers, where both possibilities are equally likely. Your expected gain from swapping envelopes, now calculated in terms of $\$b$, equals $.5 \times 1/3\$b + .5 \times -1/3\$b = 0$. But this is exactly equal to your expected gain from keeping your envelope. Thus, you should be indifferent between swapping and keeping your envelope.

The conclusion of this argument, it appears, is as it should pretheoretically be. However, merely noting that there is another way to calculate your expected gain, one that avoids the above paradoxical conclusion, is insufficient to solve the paradox: we still need some reason for thinking this alternative method is preferable to the earlier one. It might be said that we already have such a reason, namely, that the first method leads to paradox, while the second does not. But it is has been frequently remarked in the literature that for a solution to a paradox (any paradox) to be non-*ad hoc*, it must have some independent motivation, providing an explanation of why the paradox occurs.

McGrew et al. [1997], who are also concerned with the above version of the paradox, and who also note that the paradox does not arise if expected

[1] On this version of the paradox, see for instance Albers et al. [2005].

utilities are calculated in terms of the total amount of money, are aware of the need to provide an independent motivation if it is to be made into a *requirement* that expected utilities be thus calculated. In their view, the crucial point to notice is this:

> If the (fixed) total amount [involved in the game] is, say, $3x$, then the selected envelope contains x *if it contains the smaller amount*, but it contains $2x$ if it contains the larger amount. And this means that the [calculation which leads to paradox], which assumes that the selected envelope contains *the same fixed amount* whether it is the higher or the lower envelope, is illegitimate. (p. 29)

This would be different — they say — if the amount in your envelope were determined first, and it would only then be decided, for instance by the toss of a fair coin, whether the other envelope is to contain twice or half the amount of money that is in yours; then the total amount involved in the game would depend on the amount in your envelope (and the outcome of the coin toss), and one could legitimately take the latter as fixed.

Appealing though their reasoning may seem, it could at most be a partial explanation of the paradox. For what we certainly want to know is why what normally seems to be a perfectly acceptable way of calculating one's expected gain leads in the above case to paradox. Suppose you are told that the difference between the amounts in the envelopes is $\$1$ (any other fixed amount will do as well for present concerns). Then according to McGrew et al. it would be illegitimate to call the amount in your envelope $\$a$. After all, call the total amount involved in the game $\$b$, so that your envelope contains either $.5\$b + .5\$1$ or $.5\$b - .5\$1$. Then $\$a$ refers either to the former — if it contains the larger amount — or to the latter, if it contains the smaller amount. And yet, if we calculate the expected utility of swapping in terms of $\$a$, we find that it equals $\$0$, precisely what one would pretheoretically expect. In other words, nothing goes wrong in this case, the putative illegitimacy involved notwithstanding. It seems that, at a minimum, we do not have a *complete* understanding of what engenders the two-envelope paradox as long as we are unable to account for the difference between it and the at least superficially very similar but non-paradoxical decision-making situations of the kind just described. And nothing McGrew et al. say enables us to do that.

It would thus be a mistake to think that we already have our solution to the paradox. What we do have is a demonstration that whether or not you end up with a paradoxical conclusion depends on how you represent the game-theoretic situation — and that is all I promised to show at this stage.

2. The independent motivation I aim to provide involves what is sometimes called "the symmetry requirement," according to which "[p]roblems which are essentially the same must receive essentially the same solution" (van Fraassen [1989:236). There seems to be a quite strong intuition behind this requirement, namely, that an asymmetry can only be accounted for in terms of an asymmetry. That may explain why the requirement is honored as a key methodological principle both in mathematics and in the empirical sciences.[2] In a more informal version it has wide applicability beyond those areas as well; it is for instance easily recognized as a cornerstone of our moral and legal practices (in the latter realm it is commonly known as the Principle of Equality). And, most pertinent to our concerns, Sensat [1997] presents the symmetry requirement — which he labels "invariance under isomorphism" — in the context of game-theory as an adequacy constraint on solution functions, where (roughly) a solution function for a game is a mapping that, for each player, associates with the game a set of admissible moves (like, in the two-envelope case, "swap" and "keep"). This constraint "aims at insuring that the value of the solution function does not depend on extraneous or incidental features of a game's representation" (p. 387).

To see the relevance of the symmetry requirement to the problem at hand, note that if you calculate again in terms of the amount of money in your envelope, $\$a$, your colleague's expected gain from swapping, then you will find that it is $-.25\$a$ ($= .5 \times -\$a + .5 \times .5\$a$). While, prima facie, this may seem entirely unsurprising, on closer examination there is something deeply puzzling about it. For one wonders where the asymmetry between the outcomes comes from. To be sure, there is an asymmetry between the situations you and your colleague are in: one of you holds an envelope that contains twice the amount of money that is in the other. But this is not information that goes into the expected gain calculations, which take as input only probabilities and utilities. And, by the construction of the game, the relevant probabilities and utilities are the same for you and your colleague: both of you have a 50 percent chance of holding the envelope containing the larger amount, and you may both be supposed to value the amounts in the envelopes in proportion to their monetary values. It thus would seem that calculating *your* expected gain is essentially the same problem as calculating your *colleague's*, and yet we get essentially different outcomes. Put in slightly different terms, the problem is that in the output of the expected gain calculations — as here performed — there is an asymmetry that is not traceable to any asymmetry in the information that is available as input for the calculations.

[2] See Zabell [1988], van Fraassen [1989, Ch. 10], and Kosso [2000].

No violation of symmetry arises from the alternative way of calculating expected gains I proposed, for if you calculate both your and your colleague's expected gain in terms of the total amount of money involved in the game, you will get no difference in the outcomes; the expected gain of swapping, in that case, will be $0 for both of you. Hence, it appears that the symmetry requirement provides a reason for preferring the non-paradox-engendering way of calculating your expected gain in the two-envelope situation to the paradox-engendering way.[3]

It is worth emphasizing that, for present purposes, we only need to make relatively weak assumptions about the status of the symmetry requirement. In particular, we need not assume that it expresses a metaphysically necessary truth,[4] nor even that it is a strict methodological principle that impels us to reject any theory or representation that violates it. All we need to assume is that if one theory or representation violates the principle while a second one does not do so, then, in the absence of countervailing considerations, the former is preferable (to whatever precise extent) to the latter. Naturally, in our application of it there is no question as to whether the proviso is satisfied. Quite the contrary: there is even another reason for preferring the symmetry-preserving representation, for that is non-paradoxical.

Nor need we make strong assumptions about the scope of the symmetry requirement. Even if its applicability should be rather limited, we have an especially good reason for believing that it applies in the two-envelope situation. For the symmetry requirement, or at least the intuition behind it, seems to be at the root of our verdict that, in the original description of that situation, we are facing a paradox in the first place. Why think that the same reasoning that led you to prefer swapping to not swapping would have applied had you been given the other envelope, or that — as is also often said — if you were to swap, the reasoning that led you to prefer swapping to not swapping would apply again, so that you should prefer to swap once more? The only plausible answer to these questions seems to be this: holding the other envelope in your hands is to be in essentially the same situation as you are actually in, and situations that are essentially the same should be treated in essentially the same way. Or, to focus on what may well be our most basic intuition about the puzzle, why think that the intuitively correct answer

[3] Described in these terms, the crucial difference between the present version of the two-envelope paradox and the version with looking referred to earlier is that the looking occasions a symmetry-breaking: after the looking you have obtained information relevant to the decision making that is not available to your colleague, or, if your colleague also looked, you have observed different amounts.

[4] As Leibniz appears to have thought; see Rescher [1959/60].

to the question whether you should prefer to swap is that you should be indifferent between swapping and not swapping? Again, the only plausible answer seems to be that by swapping you would come to be in the position your colleague is in now, and relative to your information state that situation is symmetrical with the one you are currently in.[5] To think the symmetry requirement applies to the two-envelope situation would thus seem a mere matter of consistency.

Does the symmetry requirement solve the paradox, then? By itself, it does not. For notice that it only requires that the expected gains for you and your colleague be *equal*, not that they both be $\$0$, so that the players would be indifferent to swapping, as we pretheoretically think they should be. And while symmetry considerations have us prefer a calculation of expected gains in terms of the aggregate amount to one in terms of the amount of money in either of the envelopes, it does not follow from the above that they have us prefer the former to any other expected gain calculation. If some other such calculation should satisfy the symmetry requirement but still lead to the conclusion that you should prefer to swap envelopes, or to the conclusion that you should prefer to keep your envelope, we would still face the two-envelope paradox.

3. However, the third step offers a very simple argument to the effect that, for the game-theoretic problem at hand, an expected gain calculation satisfies the symmetry requirement only if it is non-paradoxical.

To see this, first observe that there are just three possibilities for such calculations: according to them the expected gain is, respectively, $\$0$ for both, $\$c$ for both (for some $c > 0$), or $\$d$ for both (for some $d < 0$). But, clearly, you and your colleague are involved in a zero-sum game, that is, a game in which the total amount of money involved remains constant so that one party's gain must always equal the other party's loss (or, in the general case, the sum of the gains of all players must equal the sum of the losses of all players in each round). As a result, the second and third cases, in which you and your colleague are supposed both to win and, respectively, both to lose,

[5] In fact, the symmetry of the game seems uncontestable, and has been noted before in the literature; see, for instance, Norton [1998]. What so far appears to have gone unrecognized, however, is that it takes but two relatively obvious additional observations (registered in the other sections of the present paper) to obtain what I think is a quite elegant solution to the paradox.

can be crossed out. This leaves the first case, in which the expected gain from swapping is $0 for both parties.[6]

Institute of Philosophy
University of Leuven
E-mail: igor.douven@hiw.kuleuven.be

REFERENCES

Albers, C. J., Kooi, B. P., and Schaafsma, W. [2005] "Trying to Resolve the Two-envelope Problem," *Synthese* 145:89–109.

Kosso, P. [2000] "The Empirical Status of Symmetries in Physics," *British Journal for the Philosophy of Science* 51:81–98.

McGrew, T., Shier, D., and Silverstein, H. [1997] "The Two-envelope Paradox Resolved," *Analysis* 57:28–33.

Norton, J. [1998] "When the Sum of Our Expectations Fails Us: The Exchange Paradox," *Pacific Philosophical Quarterly* 79:34–58.

Rescher, N. [1959/60] "Choice without Preference," *Kant-Studien* 51:142–175.

Sensat, J. [1997] "Game Theory and Rational Decision," *Erkenntnis* 47:379–410.

van Fraassen, B. C. [1989] *Laws and Symmetry*, Oxford: Clarendon Press.

Zabell, S. [1988] "Symmetry and Its Discontents," in B. Skyrms and W. Harper (eds.) *Causation, Chance, and Credence*, Dordrecht: Kluwer, pp. 155–190.

[6] I am greatly indebted to Barteld Kooi and an anonymous referee for this journal for valuable comments on previous versions of this paper.