Let Y be the random variable representing the amount in the smaller envelope, with distribution F_1, on a probability space Ω.

Let X(Y) and X'(Y) represent the amounts in the chosen envelope and the other envelope respectively. It has distribution F_2, on two events $\{\mathcal{E}_1, \mathcal{E}_2\}$. We are in $\mathcal{E}_1$ when we happen to choose the smaller envelope, so X = Y and X' = 2Y. And when we are in $\mathcal{E}_2$, X = 2Y and X' = Y. The choices are equally likely i.e. $\mathbb{P}(\mathcal{E}_1) = \int_{\mathcal{E}_1} dF_2 = 1/2$ and $\mathbb{P}(\bar{\mathcal{E}_2}) = \int_{\bar{\mathcal{E}_2}} dF_2 = 1/2$.

Let V(Y) be the final value after deciding to switch or not. It has distribution F_3 on two events $\{\mathcal{S}, \bar{\mathcal{S}}\}$ representing respectively switching (V = X'), or not switching (V = X). For example if the strategy is to always switch, F_3 is deterministic with $\mathbb{P}(\mathcal{S}) = \int_{\mathcal{S}} dF_3 = 1, \mathbb{P}(\bar{\mathcal{S}}) = \int_{\bar{\mathcal{S}}} dF_3 = 0$. More generally, F_3 can depend on the specific value of X, but not on F_1.

Claim: E[V] is the same regardless of F_3.

Proof: The expected value of V over all probability measures F_1, F_2 and F_3 is:

$$E[V] = \int_{\Omega} E[V|Y] dF_1 \tag{1}$$

Now, for a given Y, we have the choice of switching or not switching i.e.

$$E[V|Y] = \int V(Y) dF_3 = \int_{\bar{\mathcal{S}}} X(Y) dF_3 + \int_{\mathcal{S}} X'(Y) dF_3$$

and the choice of envelope $\mathcal{E}_1$ and $\mathcal{E}_2$:

$$\begin{aligned} E[V|Y] &= \int_{\bar{\mathcal{S}}} \left(\int_{\mathcal{E}_1} Y dF_2 + \int_{\mathcal{E}_2} 2Y dF_2 \right) dF_3 + \int_{\mathcal{S}} \left(\int_{\mathcal{E}_1} 2Y dF_2 + \int_{\mathcal{E}_2} Y dF_2 \right) dF_3 \\ &= \int_{\mathcal{E}_1} \left(\int_{\bar{\mathcal{S}}} Y dF_3 + \int_{\mathcal{S}} 2Y dF_3 \right) dF_2 + \int_{\mathcal{E}_2} \left(\int_{\bar{\mathcal{S}}} 2Y dF_3 + \int_{\mathcal{S}} Y dF_3 \right) dF_2 \end{aligned}$$

which, since F_3 (being in $\mathcal{S}$ or $\bar{\mathcal{S}}$) is independent of F_2 (being in $\mathcal{E}_1$ or $\mathcal{E}_2$)

$$= \left(\int_{\bar{\mathcal{S}}} Y dF_3 + \int_{\mathcal{S}} 2Y dF_3 \right) \int_{\mathcal{E}_1} dF_2 + \left(\int_{\bar{\mathcal{S}}} 2Y dF_3 + \int_{\mathcal{S}} Y dF_3 \right) \int_{\mathcal{E}_2} dF_2 \tag{2}$$

Thus, plugging in the actual values for F_2

$$\begin{aligned} E[V|Y] &= \left(\int_{\bar{\mathcal{S}}} Y dF_3 + \int_{\mathcal{S}} 2Y dF_3 \right) \frac{1}{2} + \left(\int_{\bar{\mathcal{S}}} 2Y dF_3 + \int_{\mathcal{S}} Y dF_3 \right) \frac{1}{2} \\ &= \frac{1}{2} \left(\int_{\bar{\mathcal{S}}} (Y + 2Y)\, dF_3 + \int_{\mathcal{S}} (2Y + Y)\, dF_3 \right) \\ &= \frac{3}{2} Y \left(\int_{\bar{\mathcal{S}}} dF_3 + \int_{\mathcal{S}} dF_3 \right) = \frac{3}{2} Y \end{aligned}$$

Now plugging this back into (1), the overall expectation of V is

$$E[V] = \frac{3}{2} \int_{\Omega} Y dF_1 = \frac{3}{2} E[Y]$$

Thus the expected value is always the same, regardless of the switching strategy F_3, including never switching, and seeing X or not seeing X makes no difference. □

Note 2: Resolving the Paradox. Let's restate the E[V] calculation in terms of X.

$$\begin{aligned} E[V|X] &= \int_{\bar{\mathcal{S}}} X dF_3 + \int_{\mathcal{S}} \left(\int_{\mathcal{E}_1} 2X dF_2 + \int_{\mathcal{E}_2} \frac{1}{2} X dF_2 \right) dF_3 \\ &= \mathbb{P}(\bar{\mathcal{S}})X + \mathbb{P}(\mathcal{S}) \left(\int_{\mathcal{E}_1} 2X dF_2 + \int_{\mathcal{E}_2} \frac{1}{2} X dF_2 \right) \end{aligned}$$

and naively averaging it as

$$\begin{aligned} E[V] &= \int_{\Omega} E[V|X] dF_1 \\ &= \mathbb{P}(\bar{\mathcal{S}})E[X] + \mathbb{P}(\mathcal{S}) \left(\int_{\mathcal{E}_1} \int_{\Omega} 2X dF_1 dF_2 + \int_{\mathcal{E}_2} \int_{\Omega} \frac{1}{2} X dF_1 dF_2 \right) \\ &> E[X] \end{aligned}$$

Since the term in parentheses is greater than E[X], it seems like there's a net increase in expected value from switching. That is the paradox.

But the average is just incorrect, since F_1 is a distribution of Y, we cannot simply compute the expected value base of X using the distribution and the probability space of Y. Let's rewrite it in terms of Y. For each value of X in $\mathcal{E}_1$ the underlying space there is Y = X and for each value in $\mathcal{E}_2$, X is actually 2Y.

$$E[V] = \mathbb{P}(\bar{\mathcal{S}})E[X] + \mathbb{P}(\mathcal{S}) \left(\int_{\mathcal{E}_1} \int_{\Omega} 2Y dF_1 dF_2 + \int_{\mathcal{E}_2} \int_{\Omega} \frac{1}{2} 2Y dF_1 dF_2 \right)$$

Now we can factor out dF_2 as before, so we get

$$\begin{aligned} E[V] &= \mathbb{P}(\bar{\mathcal{S}})E[X] + \mathbb{P}(\mathcal{S}) \left(\frac{1}{2} \int_{\Omega} 2Y dF_1 + \frac{1}{2} \int_{\Omega} \frac{1}{2} 2Y dF_1 \right) \\ &= \mathbb{P}(\bar{\mathcal{S}})E[X] + \mathbb{P}(\mathcal{S}) \frac{3}{2} E[Y] \end{aligned}$$

And of course, E[X] the expected value of the first envelope is just $\frac{3}{2}E[Y]$ (half Y + half 2Y), so we get , as before:

$$E[V] = \frac{3}{2} E[Y]$$

Note 1: Knowing prior distribution. If the switching strategy has some knowledge of F_1, then, given X, we know if it's more likely to be Y or 2Y, so the conditional probability of being in $\mathcal{E}_1$ or $\mathcal{E}_2$ is not equal. Thus we would choose an F_3 where the probability of $\mathcal{S}$ is higher if we the probability of being in $\mathcal{E}_1$ is higher, so the step (2) in the proof where we factor out dF_2 from dF_3 is no longer valid. In fact, you *can* gain by (sometimes) switching, you have to optimize F_3 based on F_1.