Question: For the randomized response exercise, why can't we use

$$0.7 \times \frac{1}{6} + 0.3 \times \frac{5}{6} = 0.3666\ldots$$

as the answer?

Short answer: The logic here seems to be: the drug users can be divided into two categories: those who answered Yes and those who answered No; we can find the probability of each category, then add them up.

However, the problem is that the probability of a drug user who answered Yes is *not* $0.7 \times \frac{1}{6}$. The 0.7 here contains Yes's from both drug users and non-drug users. It is not immediately obvious how to extract the proportion that comes from drug users, and multiplying by $\frac{1}{6}$ is certainly not the right operation here.

Long answer: We can try to find the proportion of drug users, Pr(Drug), using the *law of total probability*:

$$\begin{aligned}\Pr(\text{Drug}) &= \Pr(\text{Drug}|\text{Yes})\Pr(\text{Yes}) + \Pr(\text{Drug}|\text{No})\Pr(\text{No}) \\ &= \Pr(\text{Drug}|\text{Yes})\, 0.7 + \Pr(\text{Drug}|\text{No})\, 0.3\end{aligned}$$

But it is not immediately clear what the term Pr(Drug|Yes) (that is, probability that someone is a drug user given that they answered Yes) equals.

Note, however, that $\Pr(\text{Yes}|\text{Drug}) = \frac{1}{6}$. So the mistake essentially comes from <u>mixing up</u> Pr(Drug|Yes) with Pr(Yes|Drug).

Remark 1: you can actually relate these two terms using *Bayes' theorem*:

$$\Pr(\text{Drug}|\text{Yes}) = \frac{\Pr(\text{Yes}|\text{Drug})\Pr(\text{Drug})}{\Pr(\text{Yes})}.$$

However, if you continue this way, then all the Pr(Drug) terms will eventually cancel out, so you won't arrive at an answer by this method.

Remark 2: the correct solution (given in the slides) can be written in terms of the law of total probability as

$$\begin{aligned}\Pr(\text{Yes}) &= \Pr(\text{Yes}|\text{Drug})\Pr(\text{Drug}) + \Pr(\text{Yes}|\text{No drug})\Pr(\text{No drug}) \\ 0.7 &= \frac{1}{6}\Pr(\text{Drug}) + \frac{5}{6}\big(1 - \Pr(\text{Drug})\big)\end{aligned}$$

from which Pr(Drug) can be solved.