Meeting 11, S07
Exercise 3: Logic problem (from “What is the name of this Book?” by Raymond Smullyan)
Since it’s near the time of Halloween, it is appropriate that we have a problem set in Transylvania, Dracula’s homeland. Assume half the inhabitants of Transylvania are human and the other half are vampires. While these kinds of inhabitants cannot be distinguished by their physical appearance, the humans always tell the truth and the vampires always lie. However, to make things more complicated, half of the inhabitants of Transylvania are also insane and deluded in their beliefs; this means that for the insane Transylvanians, all true propositions they believe to be false, and all false propositions they believe to be true. The sane Transylvanians know which propositions are true and which are false.

So there are 4 categories of Transylvanians: (i) sane humans, (ii) insane humans, (iii) sane vampires and (iv) insane vampires. Whatever a sane human says is true and whatever an insane human says is false; whatever a sane vampire says is false and whatever an insane vampire says is true. For example, a sane human will say $2 + 2$ equals $4$, but an insane human will say it does not equal $4$ (because he believes that it doesn’t); a sane vampire will say $2 + 2$ is not equal to $4$ (because he lies) and an insane vampire will say they are equal (because he believes they aren’t, and then he lies about what he believes). WHEW – is that all clear?

a. If you meet a Transylvanian who says “I am human or I am sane” which of the 4 categories is he in?

b. I met a Transylvanian whom I asked “Are you an insane vampire?” Whether he answered “yes or “no”, I knew what he was. What was he?

c. Define a reliable Transylvanian as one who makes true statements (i.e., a sane human or an insane vampire) and an unreliable Transylvanian as one who makes false statements (i.e., an insane human or a sane vampire – making these statements out of delusion or malice, respectively). Suppose you ask a Transylvanian “Are you reliable?” and he answers “yes” or “no”. Can you determine from his answer whether or not he is a vampire? Can you determine whether or not he is sane?

d. Suppose instead of the question in c, you ask “Do you believe that you are reliable?” and he gives you a “yes” or “no” answer. Now can you determine whether or not he is a vampire? Can you determine whether or not he is sane?