The Official Math Club Newsletter
Volume I, Issue I

hexy

February 20, 2018

Something about hats?

2024 Edit: This newsletter was originally composed in 2018 as a start to a Math Club newsletter for college, but it never really took off. I thought that the article itself and my solution to the problem was valuable, so I'm adapting it from it's previous context and putting it up here. Enjoy!

I do not know its origin, but this problem comes from Alan Wiggins's "A List of Interesting Problems" paper. Here is the problem:

> **Problem.** A prison warden has 3 prisoners, and wants to play a game. To play the game, the warden blindfolds each of the 3 prisoners and tells them he has 5 hats: 3 are black and 2 are white. He puts a hat on each prisoner and hides the remaining hats. Then, he puts the prisoners in a circle and removes each person's blindfold one at a time. If any of the three can deduce the color of their own hat by seeing the color of the other two hats, all of them go free. If some person guesses wrong or they tell someone else the color of their hat, then they all stay in prison. There is no penalty for not guessing at all.
> The warden takes off the blindfold of the first prisoner, but he says he can't determine the color of his hat. The warden takes off the blindfold of the prisoner 2, but he also says he can't determine the color of his hat. Before the blindfold is taken off of the last prisoner, he correctly guesses the color of his own hat! What was the color of the last prisoner's hat, and why?

To help with illustrating the problem, we will draw a picture. Here the three prisoners are seated in a circle, with the first prisoner in the upper left, the second in the upper right, and the third at the bottom. Note that each of the prisoners can see the other two's hats, but not their own.

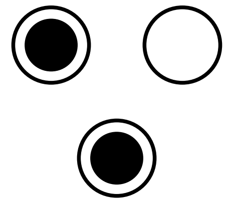

Figure 1: The three prisoners, each with white or black hats.

The problem itself seems quite bizarre. Two of the prisoners say "I don't know", and somehow that information conveys the hat color of the remaining prisoner. How does not knowing what the answer is somehow give you the answer?

To solve this problem, we will go through the description of how the game plays out one step at a time. At the first step, prisoner 1 can see the other two prisoners' hats, and declares that he doesn't know what his hat color is. In Figure 1, the first prisoner can see one black hat and one white hat. Since their hat could be either white or black without contradicting anything, he wouldn't know his hat color in this case. This actually *eliminates a possibility where the first prisoner would have known the answer immediately!*

Suppose prisoner 1 could see two white hats. Then the prisoner 1's hat can't be white: otherwise, there would be three white hats total, and the warden only has two white hats. So the prisoner 1 would be able to conclude that their hat is black, and they would have given their answer right then and there.

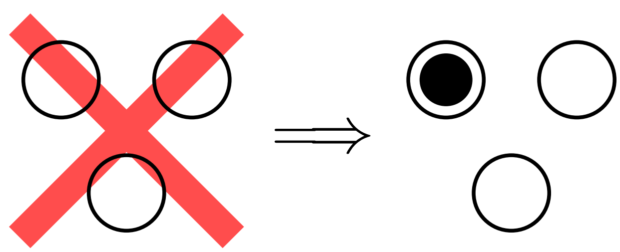

Figure 2: If a prisoner can see two white hats, they know that their hat must be black. Otherwise, there would be three white hats, and the warden doesn't have that many.

So when prisoner 1 says that they cannot tell what the answer is, what they are really saying is that they don't see two white hats. To rephrase it slightly, what the first person is saying is that they see at least one black hat. In general, when someone says "I don't know" in a puzzle like this, you should interpret it as "I can't solve the puzzle immediately, so ______ must be true".

Now, we move on to prisoner 2. They also say that they don't know what the answer is. By the same logic as before, we know that they also see at least one black hat. Knowing that the first two prisoners can see a black hat actually considerably reduces the number of possibilities! When you think about it, the only way that prisoner 3 has a white hat is if both prisoner 1 and prisoner 2 have black hats.

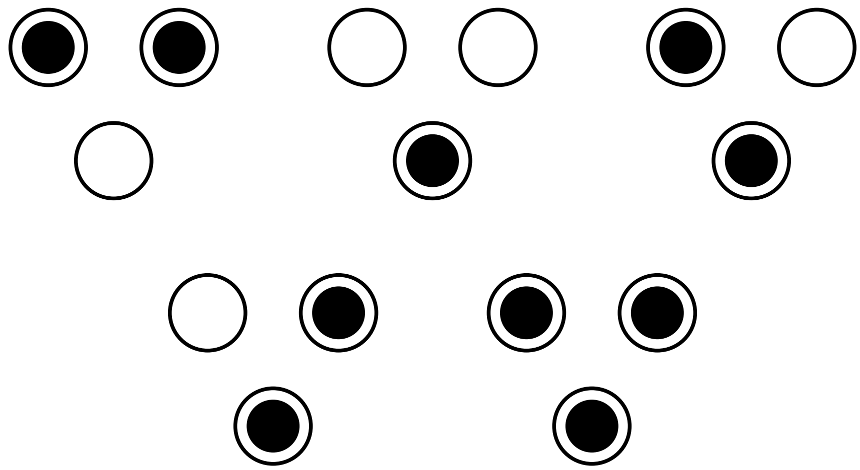

Figure 3: All possible configurations of hats where the first two prisoners see at least one black hat. Only the first configuration has the third prisoner with a white hat.

So if we can show that prisoner 3 having a white hat can't happen, then we know that prisoner 3 will have a black hat, and the reason why.

Suppose that prisoners 1 and 2 have black hats, and prisoner 3 has a white hat. Then prisoner 1 can't tell what their hat color is because they see one black hat and one white hat, so their hat could be either black or white. As we move on to prisoner 2, they can also see only one black and one white hat, but in particular prisoner 1 has the black hat. This is enough information for prisoner 2 to deduce his hat is black: if his hat was white, then prisoner 1 would see two white hats, and by the logic before prisoner 1 would have known his hat color. So if this situation really occurred, prisoner 2 would have guessed their hat color correctly, and this situation is impossible.

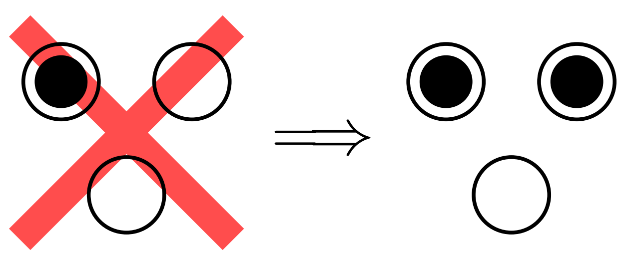

Figure 4: If prisoner 2 had a white hat, then prisoner 1 would have known he had a black hat.

By the logic we established previously, this means prisoner 3 immediately knows their hat is black, and he didn't have to see anything to figure this out. This solves the problem.

Problems

I will not be giving solutions for the moment. I wish you best of luck at attempting the problems.

1. Suppose the problem above is set up the same as before, but now the warden has 3 black hats and 3 white hats. Show that none of the prisoners can ever figure out their hat color.

2. Consider an alternative formulation of this problem: instead of trying to figure out why the third person could guess their hat color given what the first two people say, we want to instead determine the following: is there a hat configuration where the three prisoners can never guess their hat color?

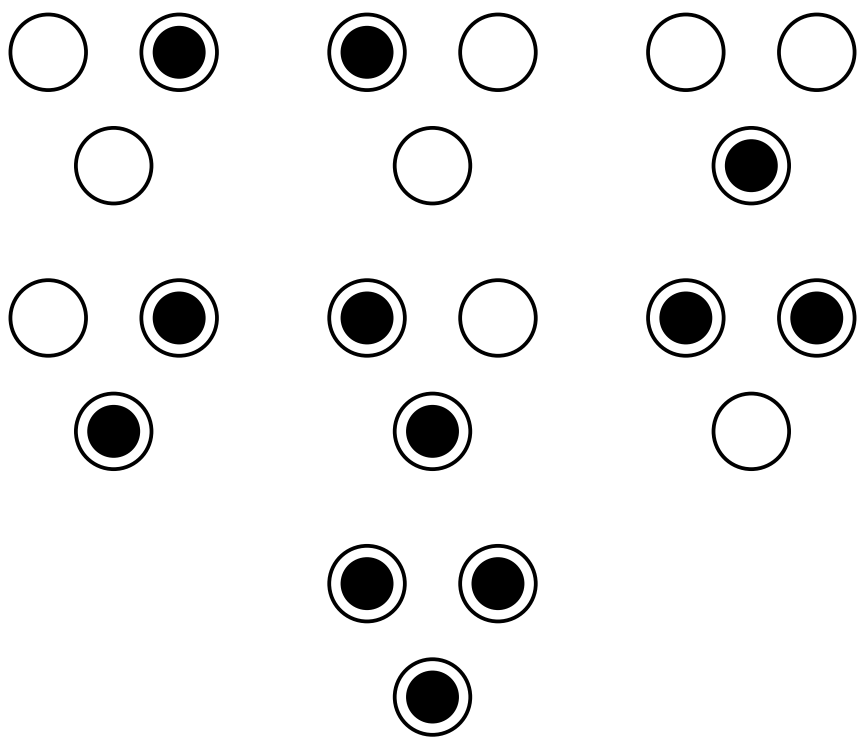

Figure 5: All possible hat configurations. There are a total of 7.

3. Consider the previous problem, but now we have 4 people, 4 black hats, and 3 white hats. Can the prisoners always figure it out or are there some situations where this fails? If it does fail, what number of black or white hats would make this work? Note that there are a total of 15 positions to consider.

4. Show that in the case of n people, n black hats, and 2 white hats (where $n \geq 3$) that the prisoners can always figure out their hat color.

5. In general, if we have n people and n black hats, what is the maximum number of white hats the warden can have so that the prisoners can always figure out their hat color?

6. Here is an entirely different problem that involves hats. It requires none of the techniques used above. We again have the same setup of the prisoners and a warden, but now there are 5 prisoners and there is a different game. The game goes like this: the warden puts either a white hat or a black hat on each of them and arranges them in a line, where each prisoner can only see the people in front of them.

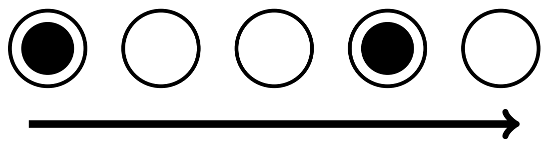

Figure 6: 5 people in a line. The person on the left can see all the people to the right of him, but the person at the far right can see no one.

The warden now has an unlimited collection of black or white hats, so there could be any combination of black or white hats on the prisoners' heads. The warden starts at the left and asks the person the color of their hat. If they guess correctly, they get to go free. Otherwise, they are sent back to prison. The warden continues in this manner all the way down the line, until everyone has been questioned. What strategy can the prisoners decide on so that as many people as possible go free, no matter what the color of their hats are?