Colored Hats and Logic Puzzles

Alex Zorn

January 21, 2013

1 Introduction

In this talk we'll discuss a collection of logic puzzles/games in which a number of people are given colored hats, and they try to guess the color of the hat on their own head. Based on the structure of the game, the players will be able to glean information about their own hat from the actions of others. Some of the games are cooperative, and the players can intentionally give other players helpful information encoded with a guess.

The analysis of these games has applications to information theory and data transmission. They're also fun puzzles to think about! In addition, these types of questions are common in interviews with companies like Google and Jane Street Capital.

Each person will be given a black and a red playing card. To simulate putting a random hat on one's head, the participants will place a playing card at random on one's forehead. In this way the students can simulate the games and get a feel for how they work.

2 Guess Your Color Competition

THE GAME: Divide into groups of three. Each player closes his or her eyes and puts on a red or black hat at random. Simultaneously all players open their eyes, and each player who sees a red hat raises his or her hand. After this, the first player to guess their hat color correctly wins (Careful- you only get one guess!)

ANALYSIS: How many distinct scenarios are there? What happens in each scenario? What is the 'trickiest' scenario?

EXPANDING: Play the same game, but in groups of four. Repeat the same analysis. What happens if you try to play it cooperatively (Try to get everyone to guess their own color correctly)? What's the main issue?

A VARIANT: Once everyone's eyes are open and hands are raised, instead of everyone guessing whenever they want, play the game in 'rounds': Have the players count '1-2-3-' and then simultaneously either guess a color or "pass". If everyone passes, play another round until somebody guesses their color correctly. What happens in this game? Can you generalize to N number of players? (This game has a lot of similarity with the 'Blue Eyes' logic puzzle).

FURTHER THOUGHT: Can you play this game with more colors? What happens? What if, instead of people raising hands to indicate what they see, the players are told beforehand that there can only be a certain number of hats a certain color (example: There are at most 2 black hats. Then if a player sees 2 black hats, they know their own must be red)?

3 Prisoners in a Line

THE GAME: In groups of four, make a line. Each person ("prisoner") puts on a hat, either red or black, so that each person can see all the colors in front of him or her. Starting from the back, each person guesses a color out loud. If you guess your own hat correctly you are saved, and if you guess incorrectly you are killed. The players can strategize beforehand.

ANALYSIS: How many people can you save?

EXPANDING: What happens if you add more people? *Remark: One variant of this puzzle specifies an infinite number of prisoners standing in a line. This variant has remarkable conclusions, but it requires a bit of set theory to analyze.*

A VARIANT: Now, prisoners are allowed to "pass". One correct guess saves everybody, while one incorrect guess dooms everybody. Also, if every prisoner passes, they all die. What is the best strategy, and what is its chance of success?

FURTHER THOUGHT: What about if there are three possible colors instead of two? Four colors? N colors?

4 Prisoners in a Circle

THE GAME: Split into pairs. Each person ("prisoner") gets a red or black hat. On the count of 3, the players simultaneously guess a color. If either of them is correct about their own hat color, both prisoners go free. If not, they are killed. The players strategize beforehand.

ANALYSIS: Can you save both people? When you solve this, try to figure it out for three people and three colors.

EXPANDING: Find a solution that is guaranteed to save the prisoners with N people and N colors.

A VARIANT: Prisoners are now allowed to "pass" or guess a color. If they all pass, or if there is one incorrect guess they all die. Otherwise, they live. Can you think of an optimal strategy for 3 people and 2 colors? *Remark: This problem, with N people and 2 colors, is closely related to Hamming Codes which are used in error detection in data transmission. There is an elegant solution where N is one less than a power of 2. If N is not of this form, the best solution in general is unknown.*

FUTHER THOUGHT: A twist on the variant allows the prisoners to have a second round if they all "pass" on the first round. How can the prisoners use this to their advantage?

5 Some Answers

Guess Your Color Competition: If all three players have red hats, nobody can immediately logically deduce his or her own color. The delay this creates should allow at least one player to realize that all three have red hats. With four players, the cases (3 red, 1 black) and (4 red) both prevent any player from deducing his or her own color immediately. Theoretically after a "short" delay, a red-hatted person should be able to identify the (3 red, 1 black) scenario, and so after a "longer" delay a player will figure out the 4 red scenario. However, the ambiguity in the time of the delay makes this a much trickier game.

In the (generalized) variant, you should find that when there are exactly one or zero red hats, the players can all guess correctly on the first round. When there are $N > 1$ red hats, the red-hatted players will all figure it out simultaneously on the $N - 1$st round.

If you add more colors the result is exactly the same, unless there are no red-hatted players. In this case, nobody will have any information about their hat color and so will never be able to guess. However, you could specify that in addition to 'raising your hand if you see red', there is a special gesture for each color (so that each player knows which colors each other player sees). The result here should not be surprising: If any color appears 0 or 1 times then those people will figure it out on the first round. Otherwise, the hat color(s) that appears the least will be guessed first, on the round numbered one less than the number of hats of that color.

Prisoners in a Line: The (quite elegant) solution is as follows: The prisoners agree to assign the number 0 to black and 1 to red. Then the prisoner in back adds up all the hat numbers in front of him and responds with the answer mod 2 (in the form of a hat color). Then the second-to-last prisoner will be able to guess correctly: He adds up all the hats in front of him, and knows that adding his own hat must give the total that the prisoner in back reported. The third-to-last can also guess correctly, since he sees all the hats in front of him and he also knows the color of the prisoner behind him, since that person just said it out loud. And so on, this strategy is guaranteed to save everybody but the person in back!

If there are infinitely many prisoners, you can also save everybody except for the person in back. However the same strategy doesn't work (because the person in back can't add up the infinitely many hats in front of him). So he needs to devise some other method of differentiating between infinite strings of hat colors. Such a method requires the "Axiom of Choice", which is an idea used when talking about infinite sets that occasionally seems to imply paradoxical results. Some mathematicians reject the Axiom of Choice, but of course I am forced to accept it.

Surprisingly, multiple colors are no more difficult than one color. The prisoners simply assign each of the N colors one of the numbers from 0 to $N-1$, and the prisoner in back reports the sum (mod N) as a hat color. Just as before, each prisoner can use this information to deduce his or her own hat color.

Returning to 2 colors, if the players are allowed to pass, there is a very nice solution: If you see all red hats in front of you, guess "black". Otherwise, pass. In this way, the last person with a black hat will guess their color correctly, and the prisoners can go free. The only way this method fails is if all hats are red, and so the chance of success is $1 - 1/2^n$ (if there are n people).

With k colors and passing allowed, a similar strategy is effective: Pick one color (say "black"), and have the prisoner guess "black" if he or she sees no black hats in front of him or her, and says "pass" otherwise. This strategy works as long as there is at least one black hat, so the chance of success is $1 - (k - 1/k)^n$. I am not sure if this is the best strategy possible.

Prisoners in a Circle: This solution is similar to the prisoners in a line. Assume there are N colors and N players. Again, assign each of the colors a number between 0 and $N-1$. Also, number the players from 0 to $N-1$. Then, the player with number i guesses the color that would make the sum of all hats equal to i mod N. So if the *actual* sum of the hats is 0, then player 0 will be correct. If it is 1 then player 1 will be correct. And so on. So exactly one of the players will be correct no matter the configuration.

Now we consider the variant with 3 people and 2 colors. Here is a nice solution: First, observe that one of two things can happen. Either all prisoners have the same color (probability 1/4), or two prisoners will have the same color and the third will have a different color (probability 3/4). So, instruct each prisoner to "pass" if he or she sees two different colors, and guess the opposite color if he or she sees two of the same colors. Then if all the prisoners have the same color hat, they will all guess incorrectly. But, if two have the same color and third is different, they will all guess correctly.

In general, with N people where $N = 2^k - 1$ is one less than a power of two, the following strategy works: Number the people from 1 to N, and convert these numbers to a binary string of k digits. We add two strings using what's called "nim-sum" or "xor", and denote by the symbol $\oplus$: For each digit, in the sum that digit is 0 if the summands match in that digit, and 1 otherwise. For example:

$$0110010 \oplus 1010111 = 1100101$$

Starting from the right, the first digits of the summands are 0 and 1 which don't match, so the first digit of the sum is 1. The second digits are both 1, they match, so the second digit of the sum is 0. And so on. An important fact is that for any number k, $k \oplus k = 0$. This is because k matches itself in every digit.

Now, recall the method from before (the N people with N hat colors), where person i guessed that the sum of all the hats was i mod N. This strategy is similar. Suppose the colors are red and black. Now let T be the sum (nim-sum) of all the numbers people with red hats. The prisoners are going to guess that T is not 0. Of course, if T is 0, they will all simultaneously be wrong and be killed. But if T is not zero the prisoners will win. Here's the strategy in more detail: If person k adds up the numbers of people with red hats that he sees, and this number is 0, he will guess red (since if he were black, T would be 0). If person k adds up the numbers of people with red hats that he sees, and this number is k, he will guess black (since if he were red, T would be $k \oplus k = 0$). If person k adds up the numbers of people with red hats that he sees, and this number is neither 0 nor k, then he passes.

Now, as mentioned before, if T is actually 0, then everybody guesses incorrectly. But, if T is equal to some number k which is not zero, then person k will guess correctly and everyone else will pass. Since T can be any number between 0 and N, the probability of failure is $1/(N+1)$, and so the probability of success is $N/(N+1)$.

EXTRA: Hamming Codes and Hat Puzzles

Suppose I want to send you a message consisting of a certain number of bits. But there is a problem: Somebody (a "demon") between us will intercept the message, and they might change one of the bits before sending it on to you. I want to build some redundancy into the message, so that if you get my message with one bit error you will be able to figure out that there is an error and then correct that error. Such a code is called 'Single Error Correcting'. For a very simple example, suppose I wanted to send you a single bit: a 0 or a 1. Instead, I'll send you a 3-bit string which is three copies of the bit I want to send you: 000 or 111. Then even if the demon changes one of the bits, you will be able to identify the mistake and correct it (and correctly interpret my message). This is called the "triple repetition code", and works for messages of any length, but it is not very efficient since it makes message three times as long!

The "Hamming Codes" are ways of creating "Single Error Correcting" messages that are more efficient. The Hamming codes gives messages consisting of $2^k - 1$ digits, where k of them are redundant and the rest $(2^k - k - 1)$ are "data". So for example when $k = 3$, you send a message that is 7 bits long that has 4 bits of data. Or when $k = 5$, your message is 31 digits long and 26 of them are data. I won't explain exactly how the Hamming codes work, but I'll try to explain the connection between the codes and the most recently discussed hat puzzle.

There is a remarkable property of the Hamming Codes: EVERY bit string (of length $2^k - 1$) is either a valid message, or one error away from being a valid message. In other words, suppose I'm using the "triple repetition code" to send you a message. Even if the demon changes one of my bits, you will NEVER receive a string like 001011, because both the first and the second blocks have errors and the demon only changed one bit. But with the hamming codes, suppose we agree that I'll send you a message of length 7. Then, if the demon is allowed to mess with one bit, ANY string is a possible string that you might receive.

Now, suppose I have $N = 2^k - 1$ people wearing either black or red hats. Then if we look at those people in a row and say black is 0 and red is 1, the hat configuration is a binary string ("message") which is N bits long. Now, each individual is able to see all of the bits but his own. If a prisoner looks at the other bits, he considers two possibilities: Either he is a 0 (black) or a 1 (red). The prisoners guess that the message is NOT a valid Hamming encoded message. Explicitly: If a prisoner concludes that him having a 0 (black) would result in a valid message, he picks red, and vise versa. If he determines that both choices (0 or 1) would result in invalid messages, he passes. This strategy messes up when the "message" IS a valid Hamming encoded message. But for every valid message there are N invalid messages (gotten by changing one of the N bits), which means the probability of success is $N/(N+1)$, the same as the optimal strategy above.

A very valuable source was E. Brown; J. Tanton, "A Dozen Hat Problems", `http://www.math.vt.edu/people/brown/doc/dozen_hats.pdf`