Guess Your Hat Color, A Brain Teaser

Owen M. Dix

Here is the riddle as I've heard it. One hundred captives are being held in one area of a Nazi concentration camp, and the Nazi warden decides to play a particularly cruel game. He gives these captives 24 hours to come up with a plan to solve this game, and their lives are on the line. In 24 hours time, each of the captives will have placed on his/her head a hat colored entirely in either red, green, or blue. Which of the three colored hats is used for a person may be entirely random. To survive this game, each person must guess the hat color on their own head.

The 100 captives will be lined up in random order in a straight, single-file line, facing the back of the person in front of them. Then the hats will be placed on their heads. They are not allowed to see the color of the hat ahead of time, and if they try to peek at their own color, they will be shot. They are allowed to talk only to say their own hat color. Any other utterances will get that person killed. They are allowed to look around but not face backwards since seeing the person behind them could easily allow for some form of communication beyond shouting out their own hat color. Leaning left or right is okay so the captives will be able to see the hats of all people in front of them, but since they cannot turn around, they won't be able to silently communicate this. Accept this as a given. After the hats get placed on their heads they can take their time to answer, but the person in back will answer first and the guards will work their way up to the front of the line, to the person who cannot see anybody's hat. If evidence is found that any of the 100 captives is cheating, trying to communicating verbally or nonverbally - even by patterns in their inflection - in any other way than shouting their own hat color out to the warden, everyone will be killed.

Clearly actually doing this would be horrible, but this does make an interesting brain teaser. Your goal is to find the strategy that, on average, would lead to the highest number of survivors, without cheating.

1 Hints

- Don't try to find a secret way to communicate information, outside of the hat color each person shouts in order. Other forms of communication are considered cheating and spoil the fun.
- Just guessing would lead to an average of $33\frac{1}{3}$ survivors out of the original 100.
- If performed flawlessly, the optimal strategy would lead to an average of $99\frac{1}{3}$ survivors.

Spoilers: the next hints are encoded, and can be read one-by-one for incremental help.

- (Read two words, skip one, repeat.) As far and as I people know, you eat will need people some math hats concepts to undo solve this juggalo problem.
- (Read two words, skip one, repeat.) All information complexity about what person hat someone too is wearing can must come with from the answer guesses of time the people ahead before them, time using what simpler they can fourth see ahead strategy of them, but and knowledge conceptualize of the concentration code.
- (Read every other word, starting with the second.) Stop assign mothers a happy number until to over each concentration hat captives color, head the most simplest will being alpha zero, beta one, gamma and delta two echo.
- (Read every other word, starting with the third.) Games form everyone tries sums hundred the easily hat math colors peek in person front shot of communicate them inflection.
- (Read every other word, starting with the fourth.) Jewish camp cannot the single first line guesser from gets are no they information blue so of they a should I guess you the an sum if of for all to hat alpha colors time in on front with of beyond them able as hear their all own in hat face. How why?
- (Read two words, skip one, repeat.) All sums take can be undone mapped into treasure the three bears numbers if sneeze you take front the remainder color of the trying sum when multiply divided by alpha three. Information shout is preserved average but you interest need to see work out headed what calculation than everyone must than make.

2 Solution

The captives define the following mapping:

$$\begin{aligned} \text{red} &= 0 \\ \text{green} &= 1 \\ \text{blue} &= 2 \end{aligned}$$

Note that the colors are ascending in number in roygbiv order, to make it easier to remember how to map the number back to the colors. Each person will have a hat value for their own head. They will do some math, get a number (either 0, 1, or 2), and map that back to the appropriate color above to yell it out as their guess. The i-th person's hat value is symbolized as h_i:

$$h_i \in \mathbb{Z}_3 = \{0, 1, 2\}$$

Here $\mathbb{Z}_3$ is the cyclic group of non-negative integers less than 3, which is just the set of numbers 0, 1, and 2. $\in$ means "is an element of". So this equation says: "the hat numbers can only take on the values 0, 1, and 2". The index, i will start from the back of the line and work towards the front. Because each person may not guess their hat correctly, we should also define g_i, the guess value for the i-th person. Clearly, if they don't guess their actual hat value (mapped back to the colors), then they will die:

$$\begin{aligned} g_i = h_i &\Rightarrow \text{live} \\ g_i \neq h_i &\Rightarrow \text{die} \end{aligned}$$

Finally, the key to surviving is to use the modulo-3 function: take the remainder of a number when you divide it by 3. This function will always return a number within the cyclic group $\mathbb{Z}_3$, so it can be mapped to one of the three colors. This is commonly symbolized with the % operator, e.g.

$$\begin{aligned} 0\%3 &= 0 \\ 1\%3 &= 1 \\ 2\%3 &= 2 \\ 3\%3 &= 0 \\ 4\%3 &= 1 \\ 5\%3 &= 2 \\ 6\%3 &= 0 \\ 7\%3 &= 1 \\ &\dots \end{aligned}$$

We will now need to define the modulo-3 of the sum of all hat values in front of person i, as:

$$s_i = \left(\sum_{j=i+1}^{100} h_j\right)_{\mathbb{Z}_3} = \left(\sum_{j=i+1}^{100} h_j\right)\%3$$

This says taking the modulo-3 is the same as doing all your math in the cyclic group $\mathbb{Z}_3$. See the Appendix for details on how to use this sort of math. Once this is defined, and we establish a rule that everyone knows, everyone not in the very back will have the information needed to calculate their own hat value then map it back to the colors. We have to first establish the rule, that everyone has to know, for what the first person in the back will do, since they have no information. The person in the very back will guess the value they calculate for $s_{i=1}$, the modulo-3 of the sum of all hat values in front of them. Notice I am starting my index numbering from 1, not 0 like computer scientists often like.

$$g_1 = s_1$$

The person in the very back will have a 1/3 chance of guessing their hat color, since they have no information about their hat color passed to them. Since they will be randomly assigned, everyone must know this is the strategy for the one in back. There are other strategies but they are variations of this, and this is the simplest. Simple is good when asking people to do so much already. This is crucial to everyone being able to guess their hat value: each person needs to know this value for everyone behind them. Here is how they do it.

Person 1 guesses the modulo-3 of the sum of the 99 people in front of them for their own hat color, as stated in the previous equation. Whether or not they are correct about this being their own hat color, they are passing information within their guess about the hats of everyone in front of them. But person 2 already knows that value for the 98 people in front of them. And now they know information that includes their own hat. They can extract it with some math. Assume all algebra is done in $\mathbb{Z}_3$. Refer to the Appendix to see how this relates to the modulo-3 function. I'll start from $i = 1$, writing the equation for the first persons guess, expand the sum of the hats in front of him/her, s_1, and relate it to the next person's sum to generalize the relationship for all people. Recall that $g_i = h_i$ in order to live.

$$
\begin{aligned}
&\text{Compute in } \mathbb{Z}_3 \\
g_1 = s_1 &= h_2 + h_3 + h_4 + \ldots + h_{100} \\
&= h_2 + (h_3 + h_4 + \ldots + h_{100}) = h_2 + s_2 \\
h_2 &= s_1 - s_2 = g_1 - s_2 = g_2 \\
h_i &= g_i = s_{i-1} - s_i
\end{aligned}
$$

The next to last line says person 2 guesses the result of person 1's sum minus their own. The last line generalizes it to all i but person 3 (or beyond) doesn't know person 2's sum directly, because they yell out this difference as their guess, so they need to do more math because the guess is all they can hear. Let's expand s_{i-1} back to person 1.

$$
\begin{aligned}
h_i + s_i = s_{i-1} &= s_{i-2} - h_{i-1} \\
h_i &= s_{i-2} - h_{i-1} - s_i \\
h_i &= (s_{i-3} - h_{i-2}) - h_{i-1} - s_i \\
&\text{Continue expanding until } s_{i-(i-1)} = s_1 \\
h_i &= (s_{i-(i-1)} - h_{i-(i-1)+1}) - h_{i-(i-1)+2} - \ldots - h_{i-1} - s_i \\
h_i &= -s_i + s_{i-(i-1)} - \left(h_{i-(i-1)+1} + h_{i-(i-1)+2} + \ldots + h_{i-1}\right) \\
h_i &= -s_i + s_1 - \sum_{j=2}^{i-1} h_j \\
&\text{Compute in } \mathbb{Z}_3 \\
g_i = h_i &= [0 - s_i] + g_1 - \sum_{j=2}^{i-1} g_j && \text{(If } i > 1 \text{, and } g_j = h_j \text{ for } 2 \leq j \leq i-1) \\
g_1 &= s_1 && \text{(If } i = 1\text{, } \tfrac{1}{3} \text{ chance correct)}
\end{aligned}
$$

This last line says that the first person guesses the modulus of the sum of all the hats in front of them as their own hat, which gives them a 1/3 chance of getting it correct. The second to last line says everyone else guesses their hat by taking $-s_i$, which is the same as subtracting it from zero, add the first person's guess, and subtract off the guessed hat numbers of all the people between the first person and the one right behind them. It is more aesthetically pleasing to write this equation in a different order, but this is the most practical. They can begin calculating their modulus, s_i, right away. Take the negative of this (subtract it from zero, in the $\mathbb{Z}_3$ cyclic group). Then they add the first persons guess and continue subtracting each persons guess from that result until the person right behind them is reached. The result is their guess which should be the hat on their head. This equation is true as long as all the hats before person i were guessed correctly, excluding the first person. We only need their guess, g_1, which might not be their hat. If a person makes a mistake, or a guard randomly kills someone, either way we shall see the best strategy in a future section.

2.1 Examples

Let's check to see if the strategy just derived works. I will limit the number of participants to 10, and randomly assign their hat colors. I'll be using the following equations to calculate s_i and g_i.

$$\begin{aligned}
&\text{Compute in } \mathbb{Z}_3\\
s_i &= \sum_{j=i+1}^{10} h_j\\
g_1 &= s_1\\
g_{i>1} &= [0 - s_i] + g_1 - \sum_{j=2}^{i-1} g_j
\end{aligned}$$

Compare these to h_i to see if they are successful.

i :	1	2	3	4	5	6	7	8	9	10
hats:	●	●	●	●	●	●	●	●	●	●
h_i :	2	1	0	2	2	0	0	1	0	1
s_i :	1	0	0	1	2	2	2	1	1	0
g_i :	$\cancel{1}$	1	0	2	2	0	0	1	0	1
Live?	N	Y	Y	Y	Y	Y	Y	Y	Y	Y

The top row shows the index number, i, where the left side is the back of the line facing to the right. Below this is the hat color, and the third row down is the number corresponding to each person's hat color, h_i. The equation each person needs says they all have to calculate s_i, the sum of the hat colors of all people in front of them, in $\mathbb{Z}_3$, which means then take the remainder when that number is divided by 3.

$$\begin{aligned}
s_1 &= (1+0+2+2+0+0+1+0+1)\%3 = 1\\
s_2 &= (0+2+2+0+0+1+0+1)\%3 = 0\\
s_3 &= (2+2+0+0+1+0+1)\%3 = 0\\
s_4 &= (2+0+0+1+0+1)\%3 = 1\\
s_5 &= (0+0+1+0+1)\%3 = 2\\
s_6 &= (0+1+0+1)\%3 = 2\\
s_7 &= (1+0+1)\%3 = 2\\
s_8 &= (0+1)\%3 = 1\\
s_9 &= (1)\%3 = 1\\
s_{10} &= 0
\end{aligned}$$

These are the values for each person for the fourth row, s_i, in the table above. The first person in line just guesses $g_1 = s_1$, but everyone else must calculate more. Using the last equation above, everyone must subtract their sum, s_i from 0 which cycles back around within $\mathbb{Z}_3$ (consult the Appendix).

$$\begin{aligned}
g_1 = s_1 &= 1\\
[0 - s_2] = [0-0] &= 0\\
[0 - s_3] = [0-0] &= 0\\
[0 - s_4] = [0-1] &= 2\\
[0 - s_5] = [0-2] &= 1\\
[0 - s_6] = [0-2] &= 1\\
[0 - s_7] = [0-2] &= 1\\
[0 - s_8] = [0-1] &= 2\\
[0 - s_9] = [0-1] &= 2\\
[0 - s_{10}] = [0-0] &= 0
\end{aligned}$$

Now, everyone ahead of $i = 1$ adds the first persons guess (in $\mathbb{Z}_3$, so apply the modulo-3 function).

$$\begin{aligned}
g_1 = s_1 &= 1\\
[0 - s_2] + g_1 &= 0 + 1 = 1\\
[0 - s_3] + g_1 &= 0 + 1 = 1\\
[0 - s_4] + g_1 &= 2 + 1 = 0\\
[0 - s_5] + g_1 &= 1 + 1 = 2\\
[0 - s_6] + g_1 &= 1 + 1 = 2\\
[0 - s_7] + g_1 &= 1 + 1 = 2\\
[0 - s_8] + g_1 &= 2 + 1 = 0\\
[0 - s_9] + g_1 &= 2 + 1 = 0\\
[0 - s_{10}] + g_1 &= 0 + 1 = 1
\end{aligned}$$

Now, as each person shouts out their guess, people subtract the guesses of those after person 1 and before themselves (again, in $\mathbb{Z}_3$), the result is their own guess.

$$g_1 = s_1 = 1$$

$$\begin{aligned}
\sum_{j=2}^{i-1} g_j = \sum_{j=2}^{1} g_j &= 0 \quad &\Rightarrow \quad g_2 &= [0 - s_2] + g_1 - \sum_{j=2}^{1} g_j = 1 - 0 = 1\\
\sum_{j=2}^{2} g_j &= 1 \quad &\Rightarrow \quad g_3 &= [0 - s_3] + g_1 - \sum_{j=2}^{2} g_j = 1 - 1 = 0\\
\sum_{j=2}^{3} g_j &= 1 \quad &\Rightarrow \quad g_4 &= [0 - s_4] + g_1 - \sum_{j=2}^{3} g_j = 0 - 1 = 2\\
\sum_{j=2}^{4} g_j &= 0 \quad &\Rightarrow \quad g_5 &= [0 - s_5] + g_1 - \sum_{j=2}^{4} g_j = 2 - 0 = 2\\
\sum_{j=2}^{5} g_j &= 2 \quad &\Rightarrow \quad g_6 &= [0 - s_6] + g_1 - \sum_{j=2}^{5} g_j = 2 - 2 = 0\\
\sum_{j=2}^{6} g_j &= 2 \quad &\Rightarrow \quad g_7 &= [0 - s_7] + g_1 - \sum_{j=2}^{6} g_j = 2 - 2 = 0\\
\sum_{j=2}^{7} g_j &= 2 \quad &\Rightarrow \quad g_8 &= [0 - s_8] + g_1 - \sum_{j=2}^{7} g_j = 0 - 2 = 1\\
\sum_{j=2}^{8} g_j &= 0 \quad &\Rightarrow \quad g_9 &= [0 - s_9] + g_1 - \sum_{j=2}^{8} g_j = 0 - 0 = 0\\
\sum_{j=2}^{9} g_j &= 0 \quad &\Rightarrow \quad g_{10} &= [0 - s_{10}] + g_1 - \sum_{j=2}^{9} g_j = 1 - 0 = 1
\end{aligned}$$

The resultant guesses on the right hand side are exactly what was inserted into the fifth row, g_i, in the table above. The first person in line guesses incorrectly, but everyone else lives. Remember you can tell if they survive by comparing h_i to g_i. This is strong evidence that the equation I derived and used to calculate these values is correct. Let's try another, without calculating every term, explicitly. I wrote a simple Python script to solve these.

i :	1	2	3	4	5	6	7	8	9	10
hats:	● (green)	● (blue)	● (green)	● (blue)	● (blue)	● (green)	● (blue)	● (blue)	● (red)	● (green)
h_i :	1	2	1	2	2	1	2	2	0	1
s_i :	1	2	1	2	0	2	0	1	1	0
g_i :	1	2	1	2	2	1	2	2	0	1
Live?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

In this particular example, everyone lives, even the first person in line. Let's try one last randomly generated example.

$i:$	1	2	3	4	5	6	7	8	9	10
hats:	●	●	●	●	●	●	●	●	●	●
$h_i:$	2	1	1	0	1	0	2	1	0	0
$s_i:$	0	2	1	1	0	0	1	0	0	0
$g_i:$	$\emptyset$	1	1	0	1	0	2	1	0	0
Live?	N	Y	Y	Y	Y	Y	Y	Y	Y	Y

Again, the first person in line did not survive, but everyone else did. It appears the calculation is correct. It would yield $99\frac{1}{3}$ of the 100 people alive, assuming everything is done correctly. But what if someone makes a mistake or gets killed arbitrarily?

2.2 Mistake or Malevolent Caprice

What if someone made a mistake, or perhaps if the warden randomly orders people to be killed. In reality, if the warden doesn't say something, you wouldn't really know which is which if someone got killed. The people in back would know, but they can't communicate that to the people in front. Even if the warden does say something, you don't know if they're telling the truth. This could be a good application for probability theory.

If someone other than in the very back gets shot, the next person could theoretically choose to pursue one of two general approaches:

- Start over from the beginning, as if they were the one in the back, giving them a 1/3 chance of surviving.
- Try to take information from the event, the previous person getting shot, to improve their chance of surviving.

In reality, considering this is a thought experiment to see which choice is best. What people do needs to be agreed upon by everyone in the 24 hours prior to the line up. Otherwise, the people in the front might not know how to proceed.

Let's investigate the second choice, to see if someone could pull some information from the event, that someone got shot. If it was a simple error on the participant's part, then it is one of the other two hats. The next person could then assume one and perform their calculation as if their guess about the previous hat was correct. They would have a 50% chance of surviving, conditioned on it being just a mistake.

But it might not just be a mistake. If it is a malicious whim of the warden or guard, then the hat they called out was correct. If that was true, the next person would have a 100% chance of getting their hat correct (conditioned on them not making a mistake, too). Which one is more likely? Is a participant making a mistake more likely than the Nazi guard being maliciously capricious? There's a decent chance of someone making a mistake, with how difficult the calculation is. But these are Nazi's. They could very conceivably do this.

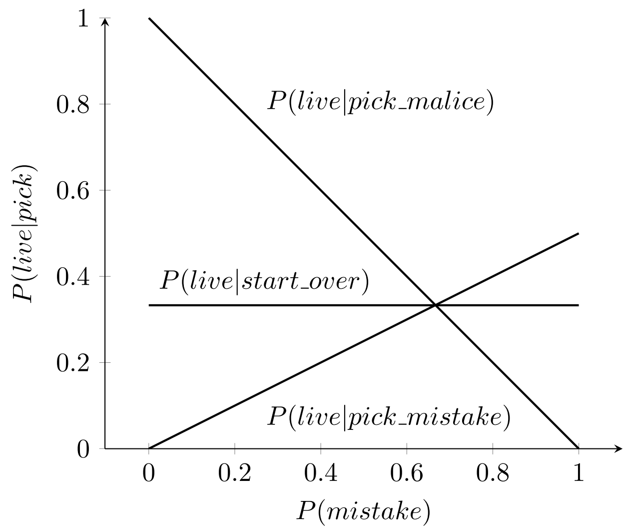
Let's assume maximum ignorance and say there's a 50% chance the shooting was caused by a mistake and a 50% chance it was malice by the guard or warden. We need to pick a method: assume it is a mistake and pick one of the other two hats, or assume it is malice and choose your hat as if the previous person got their hat correct.

$$
\begin{aligned}
P(live|pick\_mistake) =& P(live|mistake\,and\,pick\_mistake)P(mistake) \\
& + P(live|malice\,and\,pick\_mistake)P(malice) \\
& = 0.5 \cdot 0.5 + 0.0 \cdot 0.5 \\
P(live|pick\_mistake) =& 0.25 \\
P(live|pick\_malice) =& P(live|mistake\,and\,pick\_malice)P(mistake) \\
& + P(live|malice\,and\,pick\_malice)P(malice) \\
=& 0.0 \cdot 0.5 + 1.0 \cdot 0.5 \\
P(live|pick\_malice) =& 0.5
\end{aligned}
$$

This uses conditional and marginal probabilities, and note that if what you choose (*pick_mistake* or *pick_malice*) does not match what the actual cause of the shooting was (*mistake* or *malice*), then there is a zero chance of living, because they lead you to shout the wrong hat color. Here, it seems that assuming it was malice causing the shot is the best choice. But a 50% chance may not be the real odds of the shot being caused by a mistake. What is the cutoff that would maximize the chance of survival for each method? Let $p = P(mistake)$ be a variable, we optimize for, and $(1-p) = P(malice)$.

$$
\begin{aligned}
P(live|pick\_mistake) =& P(live|mistake\, and\, pick\_mistake)p \\
&+ P(live|malice\, and\, pick\_mistake)(1-p) \\
=& 0.5 \cdot p + 0.0 \cdot (1-p) \\
P(live|pick\_mistake) =& 0.5p \\
P(live|pick\_malice) =& P(live|mistake\, and\, pick\_malice)p \\
&+ P(live|malice\, and\, pick\_malice)(1-p) \\
=& 0.0 \cdot p + 1.0 \cdot (1-p) \\
P(live|pick\_malice) =& 1-p
\end{aligned}
$$

If there's a 67% chance that someone being shot is caused by a mistake on that person's part, then the chance of living by assuming it was a mistake drops to 33%, matching the odds of living by just starting over as if you're in the back and guessing. In this same event, the odds of living by guessing it was malice drops to 33% as well. The following graph shows the different choices. The optimum solution, the one that gets you the best chance of living, depends on how likely it is a shooting (after the first person) is caused by a mistake.



Choosing that the warden or guard was malicious gives the best chance of survival, for the widest range of $P(mistake)$. Only if this is greater than a 67% chance does it become best to assume they made a mistake. It is never best to start over in the event that someone other than the first person was shot. It is best just to use the guess of the previous person as if it was correct, unless after multiple test runs, you find a very high error rate.

Imagine the worst case scenario, many people get killed in a row despite all agreeing to assume that malice is causing everyone to get shot. It seems this is consistent with extreme malice (plausible for Nazi's) and extremely bad math (less likely). It still seems that assuming malice is best, but then again, you're probably going to get shot anyway, so you might as well try to make a run for it if many people in a row start to get shot. How many in a row need to get shot for this to be best? It depends on the odds of escape, which I would guess is really unlikely. So it would need to be several, especially since you running will likely result in many others getting killed. But more precisely how many, how likely it is to escape, is beyond the scope of this brain teaser.

2.3 Appendix: Basic Properties of $\mathbb{Z}_n$ and the Modulus Function

Incrementing through the cyclic group $\mathbb{Z}_n$ can be visualized by the following example diagram:

$\mathbb{Z}_3$:

Doing addition and subtraction in $\mathbb{Z}_n$ is equivalent to doing the operations in the ordinary integers, $\mathbb{Z}$, followed by taking the remainder when dividing by n, which means applying the modulo-n function to the result of the addition or subtraction. There is one caveat, which is dealing with negative numbers. I'll get to that. The following shows this equivalence relationship I just mentioned.

$$(a+b+c)_{\mathbb{Z}_n} = (a+b+c)_{\mathbb{Z}}\%n \qquad (a,b,c \geq 0)$$

Cycling through the integers in the diagram above, argues you can apply the modulus function repeatedly and your answer will always remain in $\mathbb{Z}_n$, giving you one unique answer. This has to do with the group structure of $\mathbb{Z}_n$.

$$(a+b+c+d)\%n = (((a+b)\%n+c)\%n+d)\%n$$

$$\begin{aligned}
(2+2+1+2)\%3 &= (((2+2)\%3+1)\%3+2)\%3 && \text{(Example: } n=3)\\
7\%3 &= ((4\%3+1)\%3+2)\%3\\
1 &= ((1+1)\%3+2)\%3\\
1 &= (2\%3+2)\%3\\
1 &= (2+2)\%3\\
1 &= 4\%3\\
1 &= 1
\end{aligned}$$

Negative numbers in $\mathbb{Z}_n$ wrap around to n, as you would expect by looking at the diagram at the beginning of this Appendix. The following demonstrates this by showing a negative number is just 0 minus its absolute value.

$$\begin{aligned}
-m &= 0-m\\
0 &= (pn)\%n && \text{(Let } p \in \mathbb{Z} \text{ and } pn > m)\\
-m_{\mathbb{Z}_n} &= (0-m)\%n\\
&= ((pn)\%n-m)\%n\\
&= (pn-m)\%n
\end{aligned}$$

$$\begin{aligned}
-1_{\mathbb{Z}_3} &= (0-1)\%3 && \text{(Example: } n=3)\\
&= (3\%3-1)\%3\\
&= (3-1)\%3\\
&= 2\%3\\
&= 2
\end{aligned}$$

These relationships are enough to understand algebra in $\mathbb{Z}_n$.