Mathellaneous

Norman Do

Communicating with eyes, hats and light bulbs

In his infamous book entitled *The Selfish Gene* [2], evolutionary theorist Richard Dawkins coined the increasingly popular buzzword "meme". A meme is a self-replicating unit of cultural information which is transmitted from one mind to another. As explained by Dawkins himself,

> "Examples of memes are tunes, ideas, catch-phrases, clothes fashions, ways of making pots or of building arches. Just as genes propagate themselves in the gene pool by leaping from body to body via sperms or eggs, so memes propagate themselves in the meme pool by leaping from brain to brain. . . "

Among the mathematical cognoscenti, a perfect example of a meme is a mathematical puzzle. Puzzles which are interesting and enjoyable have a habit of being continually passed around from one person to another while puzzles which are poorly contrived and uninspiring tend to die out. This highlights the parallel, originally intended by Dawkins, between memes and genes. A particularly good source of memetically successful mathematical puzzles is the topic of communication and information. For some reason, such puzzles tend to tickle the fancy of both professional mathematicians and laypeople alike. In this article, we discuss three puzzles involving the concept of communication which have become popular in the recent past. The reader is strongly recommended, if they have not come across the puzzles before, to invest some time in them before proceeding to the solutions.

Blue Eyes, Brown Eyes

> On the small island of Oculazura is a volcano as well as a tribe of 100 completely logical people. It is a well-known fact about the Oculazurans that they each have blue eyes or brown eyes. However, since there are no mirrors or reflective surfaces on the island, none of them knows the colour of their own eyes, even though they can see the colour of everybody else's. In fact, according to their strict religion, any person who is able to determine the colour of their own eyes must sacrifice themselves to the gods by jumping into the volcano at midnight on the night of their discovery.
> One day, a visiting tourist addresses the entire tribe and announces to them that at least one of the islanders has blue eyes. What happens next?

What happens next. The fate of the tribe can be determined most easily in the case that there is precisely one Oculazuran with blue eyes. Supposing that you were this person, you would already have deduced that the number of people with blue eyes on the island is either 0 or 1, depending on the colour of your own eyes. Thus, the tourist's announcement is enough for you to deduce that you must be the only person on the island with blue eyes. So on the first night, you must take the volcanic plunge, never to be seen again.

Suppose now that you are one of exactly two people on the island with blue eyes. This time, you can easily deduce that the number of people with blue eyes on the island is either 1 or 2, depending on the colour of your own eyes. If there was only one blue-eyed person on the island, then they would not live to see the second day. Therefore, when you awake the next morning to find that the tribe is intact, it dawns upon you that you must be one of exactly two people with blue eyes. Since your fellow blue-eyed islander will have reasoned in an entirely analogous manner, the two of you must sacrifice yourselves to the Oculazuran gods on the second night.

At this stage, we find ourselves in the midst of a full-fledged inductive argument. Let us now work with the hypothesis that if there are precisely k blue-eyed people, then they will commit suicide on the kth night. Now if you are one of exactly $k+1$ blue-eyed people on the island, then it is apparent to you that there are either k or $k+1$ blue-eyed people on the island, depending on the colour of your own eyes. And if you awake to see the entire tribe alive and well on day $k+1$, then you will know by the inductive hypothesis that you are one of exactly $k+1$ blue-eyed people on the island. Since the remaining blue-eyed people will have reached the same conclusion, all of the $k+1$ blue-eyed people must commit suicide on the $k+1$st night.

This beautiful inductive reasoning proves that for all n, if there are n blue-eyed people on the island, then they will plunge to their deaths on the nth night. And what about the fate of the brown-eyed people? Well, they will awake on day $n+1$ to find that n of their tribe have disappeared. It follows that those missing were the only blue-eyed people, so the remaining islanders will all know that they have brown eyes and follow their comrades into the volcano on the very next night.

A Twist in the Tale. The problem is surprising due to the fact that each islander can work out his or her eye colour from the tourist's seemingly innocent, though eventually genocidal, remark. However, the true twist in this sorry tale arises when we consider the following question.

> What information did the tourist give the Oculazurans to enable them to deduce their own eye colours?

For if we consider the case that there is more than one blue-eyed person on the island, then each of the Oculazurans must be able to see at least one pair of blue eyes. Hence, every single Oculazuran was already aware, prior to the tourist's announcement, that there existed at least one blue-eyed person on the island. So is it not true then that the tourist has only told the Oculazurans what they already knew? Does it not follow that the tourist's presence on the island was unnecessary for the mass suicide to occur? And if so, when would it have occurred? The reader is strongly urged to ponder these questions before continuing the article.

It turns out that the tourist is indeed the catalyst for the genocide. To determine what information is actually being passed to the Oculazurans that they already do not know, consider the following. What would have happened had the tourist, rather than making his announcement to the entire tribe, instead whispered the information into each person's ear in a clandestine manner? It is easy to see that no deaths would have ensued, and the islanders would have lived happily ever after (at least until the next tourist arrived on the island). The crucial fact is that not only does everyone know that at least one islander has blue eyes, but they also know that every other islander knows this fact, they know that everyone else knows that they know, and so on ad infinitum. It is this information that leads to the downfall of the poor Oculazurans.

If a group of agents all know P, then they are said to have first order knowledge. If they all know P and they know that they all know P, then they are said to have second order knowledge, and so forth. If they know P to all orders, then the proposition P is said to be common knowledge. The puzzle of the Oculazurans is an archetypal demonstration of how first order knowledge and common knowledge are manifestly different. For more information on common knowledge, the reader is directed to the relevant entry in the Stanford Encyclopedia of Philosophy [5].

Guessing Game. The following little-known problem has a similar flavour to the previous puzzle. It was told to me many years ago and, to this day, I still find it astonishing.

> Alice and Bob each choose a positive integer and reveal their numbers to Xander, but not to each other. Xander then writes two positive integers on a blackboard and tells Alice and Bob that one of them is the sum of their chosen numbers. He then alternately asks Alice and Bob, "Do you know the other person's number?" until one of them answers, "Yes". Prove that after finitely many questions, one of them will know the other player's number. (Of course, it is necessary to assume that Alice and Bob are completely logical as well as completely honest!)

Red Hats, Blue Hats

> With equal probability, a red hat or a blue hat is placed on the heads of each of three players. Each player can see the other two hats but is not able to see their own. When the signal is given, they must simultaneously guess the colour of their own hat or pass. The team will win the grand prize if and only if at least one player has guessed correctly and no players have guessed incorrectly.
> What is the optimal strategy for the team to maximise their probability of winning?

Beating the odds. Note that if everyone decides to guess randomly, then the probability of winning is only $\frac{1}{8}$. On the other hand, if one person is elected to guess randomly, then the probability of winning is increased to $\frac{1}{2}$. At first glance, it seems that this may indeed be the optimal strategy. Indeed, how can any one player gain useful information by seeing the colours of the other hats, which are chosen independently from their own?

Consider the following strategy: if a person sees two hats of different colours, they pass, and if a person sees two hats of the same colour, they guess the opposite colour. The following table shows the outcomes for the eight possible distributions of hat colours amongst the three players.

Hats			Guesses			Result
R	R	R	B	B	B	lose
R	R	B	-	-	B	win
R	B	R	-	B	-	win
R	B	B	R	-	-	win
B	R	R	B	-	-	win
B	R	B	-	R	-	win
B	B	R	-	-	R	win
B	B	B	R	R	R	lose

Therefore, we can see that this strategy allows the team to win six times out of eight — that is, with a probability of $\frac{3}{4}$. This surprising result seems to arise from the fact that whenever a player guesses incorrectly, all players conspire to guess incorrectly. On the other hand, whenever a player guesses correctly, they are alone in doing so. In fact, we can turn these observations into an argument to prove that the strategy is optimal.

Let us assume that the team adopts a deterministic strategy; in other words, one where each player's decision is determined uniquely by the hats of the other players. We can do this since any strategy which involves some sort of randomization cannot do better than a deterministic one[1]. Now suppose that the game was played eight times in succession, once for each possible distribution of hat colours. We can form a table similar to the one above to represent the guesses and outcomes. The crucial observation is that the number of correct guesses and incorrect guesses made in all eight games must be equal to the same integer, call it G. This follows since for any set of hats that a player can see, there are two possible games that can arise — one where the player has a blue hat and one where the player has a red hat.

If the team wins W times out of eight, then we must have

$$W \leq G,$$

since at least one correct guess contributes to each win. Furthermore, if the team loses L times out of eight, then we must have

$$G \leq 3L,$$

since at most three incorrect guesses can contribute to each loss. Concatenating these inequalities yields $W \leq 3L$ and with the further condition that $W + L = 8$, we obtain $W \leq 6$. Thus, the probability of winning is $\frac{W}{8} \leq \frac{3}{4}$, as required.

More players, more hats and some coding theory. Having solved the problem for three players, let us pose the question in more generality, for n players. In fact, we can perform a similar analysis to the three hat case to obtain an upper bound for the probability of winning. This time, we play the game 2^n times and let W, L and G correspond to the number of wins, the number of losses, and the number of correct (or incorrect) guesses made. By analogous reasoning, we obtain

$$W \leq G \quad \text{and} \quad G \leq nL.$$

This yields $W \leq nL$ and with the further condition that $W + L = 2^n$, we obtain $W \leq \frac{n2^n}{n+1}$. Therefore, the probability of winning $\frac{W}{2^n}$ is bounded above by $\frac{n}{n+1}$.

Whenever one obtains an upper bound, the next step is invariably to see if it can be achieved. However, in this case, we can hope for no such luck in general. For W is an integer, so this particular bound can only be achieved if $n + 1 \mid n2^n$ which implies that $n = 2^k - 1$ for some positive integer k. And in this case, the upper bound can be achieved, using a cunning strategy whose idea stems from coding theory.

---

[1] The fact that a randomized strategy cannot outperform a deterministic strategy may seem intuitively plausible to some, and downright obvious to others. A quick sketch of a proof requires one to realise that a randomized strategy is simply a convex combination of all of the possible deterministic strategies. The probability of winning is a linear function on the convex polyhedron corresponding to the set of randomized strategies. Therefore, by the standard optimisation techniques, it must achieve its maximum at a vertex of the polyhedron, which corresponds to a deterministic strategy.

For simplicity, we will describe the strategy for the case of $2^4 - 1 = 15$ players. Assign a 4-digit non-zero nimber[2] to each of the players — that is, $0001, 0010, 0011, \dots, 1111$. Each player then adds up the nimbers corresponding to all of the people they can see wearing a red hat. If the sum is 0000, then the player must guess that their hat is red. If the sum is equal to the number assigned to them, then the player must guess that their hat is blue. If the sum is equal to neither, then the player simply passes.

Let us see why such a scheme should be optimal. Suppose that the sum of the nimbers corresponding to all of the people with red hats happens to be 0000. Then all of the people with blue hats will guess that their hat is red, while all of the people with red hats will guess that their hat is blue. So in this case, everyone guesses incorrectly, as desired. On the other hand, if the sum is for example 1011, then the only person who will guess their hat colour is the person whose corresponding nimber is 1011 and they will guess correctly. So the probability of winning is the probability that the sum of the nimbers corresponding to people with red hats is non-zero. And this is simply $\frac{15}{16}$, as one might expect, given that there are 15 non-zero nimbers from the set of 16 possible.

Surprisingly, the optimal strategy has only been determined when the number of players is less than nine, one less than a power of two, or a power of two itself. In the case that the number of players is 2^k, the optimality is obtained by forgetting one of the players and using the strategy discussed above for $2^k - 1$ players. Interestingly enough, the optimal probability of winning, although not an increasing function in the number of players, approaches 1 as the number of players approaches infinity.

The astute reader may have noticed that the strategy discussed above is not only useful in winning hat-based games, but is also useful in constructing error-correcting codes. So rather than being just a frivolous mathematical puzzle, the problem is of great interest to coding theorists as evidenced by the article [4]. In fact, the problem evolved out of the PhD thesis of Todd Ebert entitled *Applications of recursive operators to randomness and complexity*, which ostensibly has nothing to do with hats whatsoever. As stated by Buhler in his article [1] from the Mathematical Intelligencer, "It is remarkable that a purely recreational problem comes so close to the research frontier."

Another hat trick. Consider the following variation on the red and blue hats theme.

> With equal probability, a red hat or a blue hat is placed on the heads of each of n players. The players are arranged in a line so that each player can only see the colours of the hats of in front of him or her. Each player must guess the colour of his or her hat. However, the guesses are not made simultaneously, but sequentially, from the back of the line to the front.
>
> If the team has a chance to collaborate beforehand on a strategy, how many players can be guaranteed to survive?
>
> Can you solve the problem if the hats can be any one of k different colours?

[2] No, this is not a typographical error! Nimbers are simply binary numbers where the additions are performed without any carries. For example, we have $1101 + 1110 = 0011$. Their name derives from the amazing fact that they appear in the analysis of the game Nim.

One Hundred Prisoners and a Light Bulb

Standing before a crowd of one hundred prisoners is the prison warden who addresses them as follows.

> "In one hour, you will each be put into solitary confinement. However, every night from now on, I will choose one prisoner at random and take them to a room where there is a light bulb connected to a switch. That prisoner may see whether the light bulb is on or off and decide whether or not to toggle its state. If at some point, you enter the room and are completely certain that all 100 prisoners have previously visited the room, then you should tell me. For if you are correct, I will set all of you free. And if you are incorrect, then you will all be put to death."

Can you devise a plan that the prisoners can use so that eventually they will all be set free?

In the event that the prison warden happens to exclude one of the prisoners from visiting the room, there is obviously no way for the prisoners to escape. However, we will assume that the prisoners are chosen uniformly at random, so that this occurs with probability 0. In particular, we will consider a winning strategy to be one which leads the prisoners to freedom with probability 1. The first response of many people to this problem is to convince themselves that a winning strategy is clearly impossible. How on earth could the prisoners count to 100 on a one bit counter? Perhaps the threat of lifetime incarceration would help to change their minds!

Strategy 1. The following strategy is perhaps the simplest to determine as well as to implement.

- The prisoners divide the time of their incarceration into blocks of 100 days.
- A prisoner entering the room on the first day of a block must turn the light bulb on.
- Every other prisoner who enters the room for the first time in a block must leave the light bulb on.
- A prisoner entering the room for the second time in a block must turn the light bulb off.
- If a prisoner enters the room for the first time in a block on the last day and sees that the light bulb is on, then he should tell the warden that he is certain that every prisoner has visited the room.

Strategy 2. When confronted with this problem, many people remain stuck in the mindset that each of the prisoners must follow the same algorithm. Of course, this is a self-imposed caveat rather than one given in the problem and doing away with it can lead to elegant solutions such as the following.

- One of the prisoners is designated the "counter" while the rest of them are considered to be "ordinary".
- The first prisoner to enter the room makes sure that the light bulb is off when he leaves the room.
- When an ordinary prisoner visits the room for the first time and finds the light bulb off, he turns it on.
- When an ordinary prisoner visits the room for a second or later time and finds the light bulb off, he leaves it off.
- When an ordinary prisoner visits the room to find the light bulb on, he leaves it on.
- When the counter visits the room to find the light bulb off, he leaves it off.

- When the counter visits the room to find the light bulb on, he turns it off and adds one to his tally.
- When the counter has reached a tally of 99, he should tell the warden that he is certain that every prisoner has visited the room.

What is the best strategy? Now that the prisoners have two strategies at hand, which one should they adopt? It seems likely that they would opt for the one which minimises the expected number of days till their release. Given that this is the case, let us consider how long that might be. The problem with Strategy 1 is that the 100 prisoners must all enter the room in one of the blocks of 100 days. Of course, this will eventually happen with probability 1, but it seems that it might take a considerably long time. Without entering into the details of the probabilistic analysis, the expected number of days for Strategy 1 is given by the expression

$$\frac{100^{101}}{100!} \approx 1.072 \times 10^{44}.$$

This is of the order of 10^{41} years, a phenomenally gargantuan length of time which far outweighs the currently adopted age of the universe. Only the purest of mathematicians amongst the prisoners might content themselves with the knowledge that they have a theoretical solution, despite the fact that the prisoners, the warden, perhaps even the universe, will probably not be around for the duration of the plan. So the prisoners will have to make do with Strategy 2. In this case, the expected number of days until release is

$$100 \times \left(\frac{2}{1} + \frac{3}{2} + \frac{4}{3} + \frac{5}{4} + \cdots + \frac{100}{99}\right) \approx 10418.$$

This is only about 29 years and probably well within the lifetime of many of the prisoners. However, is there a better strategy and, if so, how much better can we do?

The best efforts so far, rather than just having one counter, use pyramid schemes whereby everyone is involved in the process of counting. Such strategies can take as little as approximately 3900 days, or 11 years, for success although it is not known how far this is from optimal. In this case, a probabilistic analysis is harder to perform, so computer simulation is necessary to obtain this estimation. We can certainly obtain a lower bound, since the expected number of days until every prisoner has visited the room is actually

$$100 \log(100) \approx 461.$$

Variations on a theme. There are many variations on the prisoners and light bulb theme, some of which are given below. Amazingly, the prisoners can escape in all of the following circumstances. For more information on the problem, as well as more variations, and more spoilers, the reader is encouraged to consult the excellent article [3].

Everybody knows
The prisoners are all freed if and only if they have all told the warden that every prisoner has visited the room.

Eavesdropping warden
While discussing their strategy, the prisoners realise that the warden is eavesdropping on their conversation. A concerned prisoner decides to ask, "How do we know that you won't strategically choose which prisoner visits the room each night, rather than choosing us uniformly at random?" "Hmmm…that's a good idea, I think I'll do that. But to keep things fair, I'll allow each prisoner to visit the room infinitely often!" (Note, for example, that Strategy 1 fails to work for this variation.)

Multiple rooms
There are now $n > 1$ identical rooms, each with their own light bulb. The prisoners are taken to one of them at random, but do not know which one.

Messages
Each prisoner is given an arbitrarily long, though finite, message which they must convey via the light bulb. The prisoner who announces to the warden that every one else has visited the room must also know all 100 messages for the prisoners to be freed.

References

[1] J.P. Buhler, *Hat tricks*, Mathematical Intelligencer **24** (2002), 44–49.

[2] R. Dawkins, *The Selfish Gene*, (Oxford University Press New York 1976).

[3] P. Dehaye, D. Ford and H. Segerman, *One hundred prisoners and a lightbulb*, Mathematical Intelligencer **25** (2003), No 4, 53–61.

[4] H.W. Lenstra and G. Seroussi, *On hats and other covers*, IEEE International Symposium on Information Theory (2002), http://www.hpl.hp.com/research/info_theory/hats_extsum.pdf

[5] P. Vanderschraaf and G. Sillari, *Common knowledge*, in *The Stanford Encyclopedia of Philosophy* (Winter 2005 edition), E.N. Zalta ed., http://plato.stanford.edu/archives/win2005/entries/common-knowledge/

Department of Mathematics and Statistics, The University of Melbourne, VIC 3010
E-mail: N.Do@ms.unimelb.edu.au