Figure 1: Guess Who?

HOW MANY QUESTIONS DOES IT TAKE? - THE MATHEMATICAL THEORY BEHIND 'GUESS WHO?'

Created by: Gavriel Ferster

February 2026

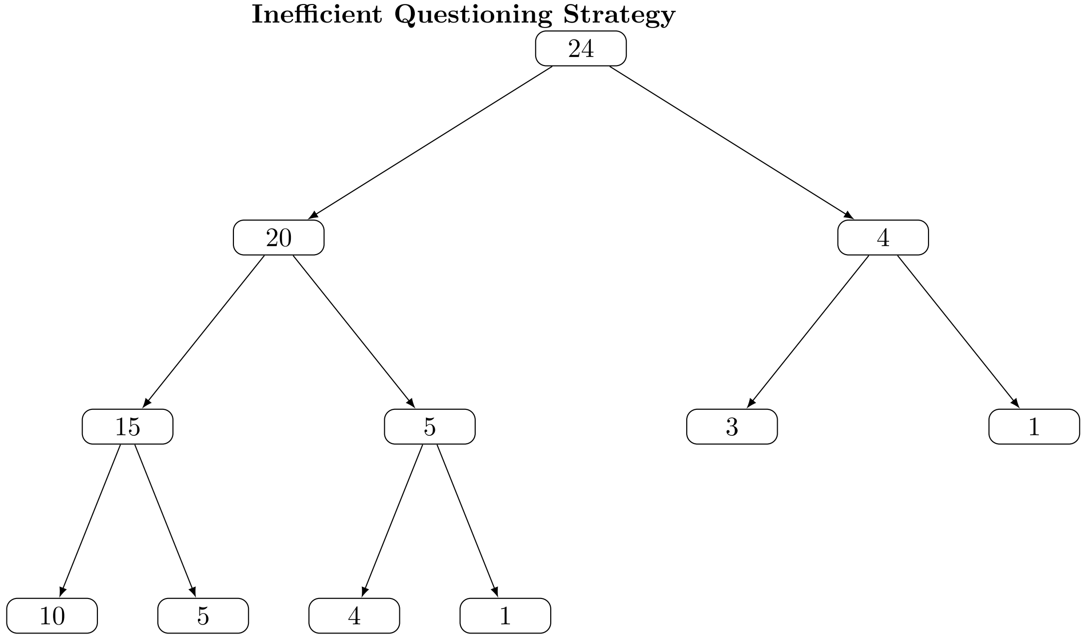
1 An introduction to the game 'Guess Who?'

I am sure most people remember 'Guess Who?' as a simple yet engaging childhood game. Played by two opponents sitting opposite each other, the game involves selecting 1 of 24 unique cartoon characters and firing a series of "yes" or "no" questions to eliminate the number of possibilities. First released in 1979 by Theo and Ora Coster, the game quickly became popular and was later published by Milton Bradley (now owned by Hasbro). The objective is straightforward: determine the opponent's chosen character before they identify yours, using logical elimination rather than random guessing. However, beneath its simple design lies an interesting mathematical problem: What is the most efficient way to narrow down 24 possibilities to just 1 using the smallest possible number of questions? or in simple terms, How many questions does it take?

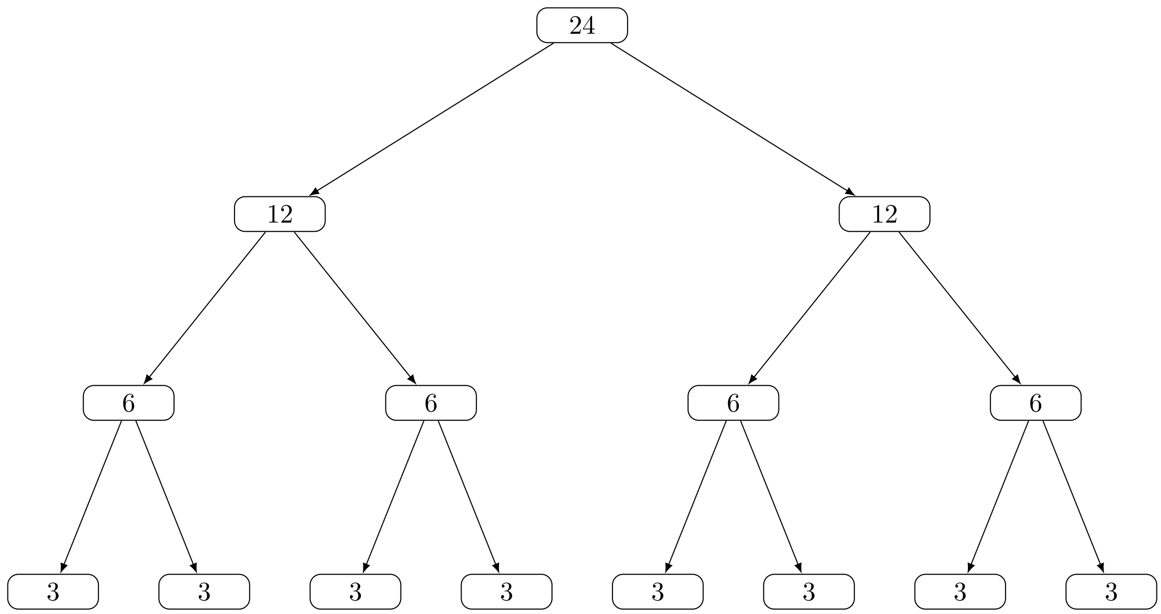
2 The Binary Decision Process of Efficiency VS Inefficiency

A binary decision process in this case scenario means that each question asked by a person has only two possible solutions "yes" or "no" where each answer divides the remaining possibilities into two smaller groups. Repeating this process gradually reduces the number of possible characters until only one remains. There are 2 scenario shown in the tree diagram below:

Inefficient Questioning Strategy



Efficient Questioning Strategy



The inefficient pathway - The first tree diagram illustrates an inefficient questioning strategy, where the questions divide the characters unevenly.

Specific questions for this to occur may include:
"Is your character wearing a hat?"
"Is your character wearing glasses?"
"Does your character have blonde hair?"
"Does your character have an E in its name?"

For example, the first question may produce a split such as

$$24 \rightarrow 20$$

or 4 If the answer leads to the larger branch, the number of remaining possibilities is still extremely large. Instead of halving the search space, the player has only removed four characters. However, in reality, the game is designed to divide the characters key features into groups of 5 such as types of hair colour and accessories.
If similar questions continue to be asked, the reduction in possibilities becomes

$$24 \rightarrow 20 \rightarrow 15 \rightarrow 10$$

where the number of remaining characters decreases much more slowly.
Mathematically, this strategy wastes information. A yes–no question can potentially reduce the uncertainty by a factor of two, but when the split is highly uneven, the information gained from the answer is much smaller.

The efficient pathway - Now notice how after each question the likelihood of guessing the opponent's character is doubled.

$$24 \to 12 \to 6 \to 3$$

until there is a 1 in 3 chance of getting it correct.
The questions asked for this to be successful are very broad. For example:

"Is your character a male?"
"Does your character's name start with a letter from A to M?"

Let us explain this mathematically: Each yes–no question has two possible answers.
After n questions, there are the fore at most

$$2^n$$

distinct sequences of answers.
To uniquely identify one character among 24, we must therefore satisfy

$$2^n \geq 24$$

Evaluating powers of two:

$$2^3 = 8, 2^4 = 16, 2^5 = 32$$

Since 32 is the first value greater than 24, we conclude:

$$n = 5$$

This proves that:

5 questions are sufficient in theory. Now lets develop this further...

3 Shannon's Source Coding Theorem

Shannon's Source Coding Theorem was developed by Claude Shannon in his famous 1948 paper: 'A Mathematical Theory of Communication'. One of Shannon's most famous results is the Source Coding Theorem, which states that any message produced by a random source can be encoded using an average number of bits at least equal to its entropy.

Linking this to 'Guess Who?', at the start of the game, each character is assumed to be equally likely of being chosen. Therefore, the probability that any particular character being selected is:

$$p_i = \frac{1}{24}$$

where p_i represents the probability of the i-th character.

To measure uncertainty, Shannon introduced the concept of entropy, defined as:

$$H = -\sum_{i=1}^{n} p_i \log_2 (p_i)$$

where:

- p_i is the probability of each outcome
- The logarithm is taken base 2 because information is measured in bits

A bit represents the information obtained from a binary decision, such as a yes-no question.

Subbing

$$p_i = \frac{1}{24}$$

gives:

$$H = -\sum_{i=1}^{24} \left(\frac{1}{24} \log_2 \left(\frac{1}{24} \right) \right)$$

As the probability is the same for every character, this becomes:

$$H = -24 \left(\frac{1}{24} \log_2 \left(\frac{1}{24} \right) \right)$$

Simplifying to:

$$H = \log_2 (24)$$

Giving us:

$$H \approx 4.58$$

bits.

Since each yes-no question can provide at most 1 bit of information, the number of questions required must satisfy the inequality:

$$n \geq 4.58$$

But questions must be an integer, so:

$$n = 5$$

This result aligns with the earlier reasoning when using the decision trees and powers of 2.
If the questions divide the possibilities as evenly as possible, uncertainty decreases rapidly until only 1 character remains. However if questions produce uneven splits, the reduction in entropy is significantly smaller.

4 The Mathematical Comparison

The amount of information gained from asking questions about the opponent's character depends on the probability of outcomes.
We can model this using the formula:

$$H = -p\log_2(p) - (1-p)\log_2(1-p)$$

Where:
• p is the probability of 1 outcome

So if a question divides the characters evenly then:

$$p = 0.5$$

Subbing p into the formula:

$$H = -0.5\log_2(0.5) - 0.5\log_2(0.5)$$

Since:

$$\log_2(0.5) = -1$$

We obtain:

$$H = 1$$

bits.

So linking back to the entropy formula and considering the uneven split such as:

$$24 \to 20, 4$$

So the probabilities are:

$$p = \frac{20}{24}, 1 - p = \frac{4}{24}$$

Subbing into the formula gives:

$$H = -\frac{20}{24}\log_2\left(\frac{20}{24}\right) - \frac{4}{24}\log_2\left(\frac{4}{24}\right)$$

Which gives:

$$H \approx 0.65$$

bits.

5 Non Uniform Probabilities

In previous analysis we assumed that every character in 'Guess Who?' is chosen with equal probability of $p = 0.5$ but now lets suppose 1 character is selected with probability:

$$p_1 = 0.2$$

and the remaining 23 characters share the remaining probability:

$$p_i = \frac{0.8}{23}$$

for

$$i = 2, ..., 24$$

The entropy formula then becomes:

$$H = -\left(0.20\log_2\left(0.20\right)\right) - 23\left(\frac{0.80}{23}\log_2\left(\frac{0.80}{23}\right)\right)$$

Giving us:

$$H \approx 4.28$$

bits.
But since $4.28 < 4.58$, the uncertainty is lower than in the uniform case. Thus the theoretical number of questions satisfies:

$$n \geq H$$

So:

$$n \geq 4.28$$

Which still implies that when rounding up (only way possible):

$$n = 5$$

questions, but this time, as H was smaller then before, the average information about the character has now decreases but only at a minimal cost.

6 Conclusion

To conclude, by modelling the game 'Guess Who?' as a process of reducing uncertainty through strategic questioning and comparing efficient vs inefficient decision pathways, its clear that asking effective questions can divide the remaining characters on the board as evenly as possible.

By using ideas from the information theory by Claude Shannon we can see that in theory only a small number of precisely 5 well chosen yes-no questions are all that is needed to identify the opponent's character and win the game. Overall, this investigation shows how a simple childhood game can be understood deeper through the power of Mathematics.