Activity 5

Twenty Guesses—*Information Theory*

Summary

How much information is there in a 1000-page book? Is there more information in a 1000-page telephone book, or in a ream of 1000 sheets of blank paper, or in Tolkien's *Lord of the Rings*? If we can measure this, we can estimate how much space is needed to store the information. For example, can you still read the following sentence?

Ths sntnce hs th vwls mssng.

You probably can, because there is not much 'information' in the vowels. This activity introduces a way of measuring information content.

Curriculum links

- ✓ Mathematics: Number Level 3 and up. Exploring number: Greater than, less than, ranges.
- ✓ Algebra Level 3 and up. Patterns and sequences
- ✓ English

Skills

- ✓ Comparing numbers and working with ranges of numbers
- ✓ Deduction
- ✓ Asking questions

Ages

- ✓ 10 and up

Materials

- ✓ No materials are required for the first activity

There is an extension activity, for which each child will need:

- ✓ Worksheet Activity: Decision trees (page 40)

Twenty Guesses

Discussion

1. Discuss with the children what they think information is.

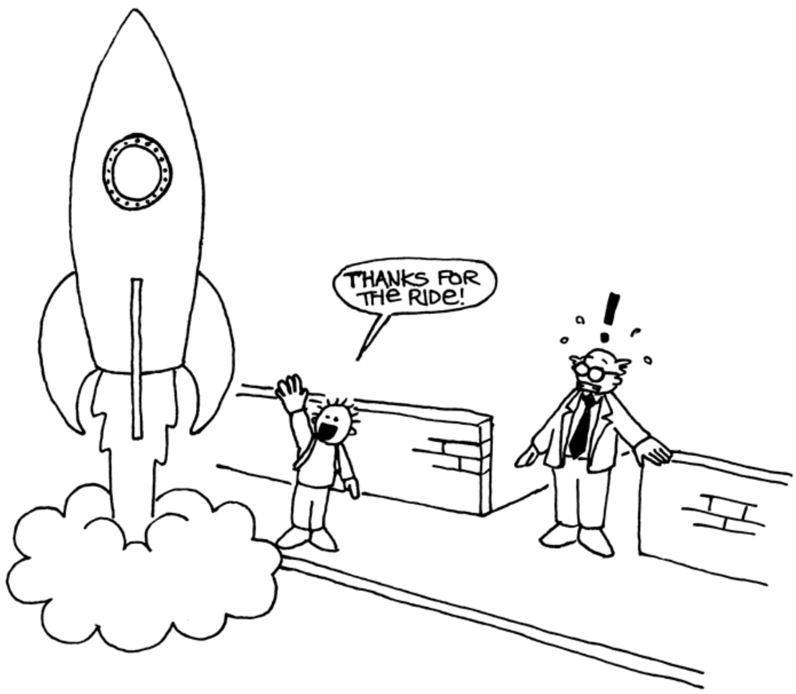
2. How could we measure how much information there would be in a book? Is the number of pages or number of words important? Can one book have more information than another? What if it is a very boring book, or a particularly interesting one? Would 400 pages of a book containing the phrase "blah, blah, blah" have more or less information than, say, the telephone directory?

   Explain that computer scientists measure information by how surprising a message (or book!) is. Telling you something that you know already—for example, when a friend who always walks to school says "I walked to school today"—doesn't give you any information, because it isn't surprising. If your friend said instead, "I got a ride to school in a helicopter today," that *would* be surprising, and would therefore tell us a lot of information.

   How can the surprise value of a message be measured?

   One way is to see how hard it is to guess the information. If your friend says, "Guess how I got to school today," and they had walked, you would probably guess right first time. It might take a few more guesses before you got to a helicopter, and even more if they had travelled by spaceship.

   The amount of information that messages contain is measured by how easy or hard they are to guess. The following game gives us some idea of this.



Twenty Questions Activity

This is an adapted game of 20 questions. Children may ask questions of a chosen child, who may only answer yes or no until the answer has been guessed. Any question may be asked, provided that the answer is strictly ‘yes’ or ‘no’.

Suggestions:

I am thinking of:

- ✓ a number between 1 and 100
- ✓ a number between 1 and 1000
- ✓ a number between 1 and 1,000,000.
- ✓ any whole number
- ✓ a sequence of 6 numbers in a pattern (appropriate to the group). Guess in order from first to last. (e.g. 2, 4, 6, 8, 10)

Count the number of questions that were asked. This is a measure of the value of the “information”.

Follow-up Discussion

What strategies did you use? Which were the best ones?

Point out that it takes just 7 guesses to find a number between 1 and 100 if you halve the range each time. For example:

Is it less than 50?	Yes.
Is it less than 25?	No.
Is it less than 37?	No.
Is it less than 43?	Yes.
Is it less than 40?	No.
Is it less than 41?	No.
It must be 42!	Yes!

Interestingly if the range is increased to 1000 it doesn’t take 10 times the effort—just three more questions are needed. Every time the range doubles you just need one more question to find the answer.

A good follow up would be to let the children play Mastermind.

Extension: How much information is there in a message?

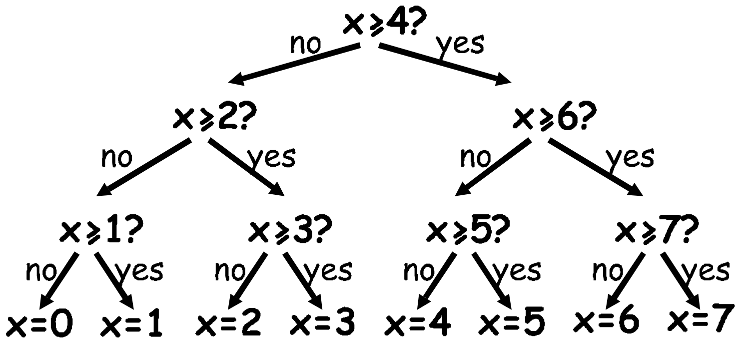
Computer scientists don’t just use guessing with numbers—they can also guess which letter is more likely to be next in a word or sentence.

Try the guessing game with a short sentence of 4–6 words. The letters must be guessed in the correct order, from first to last. Get someone to write down the letters as they are found and keep a record of how many guesses it takes to find each letter. Any questions with a yes/no answer can be used. Examples would be, “It it a *t*?” “Is it a vowel?” “Does it come before *m* in the alphabet?” A space between words also counts as a “letter” and must be guessed. Take turns and see if you can discover which parts of messages are easiest to find out.

Worksheet Activity: Decision Trees

If you already know the strategy for asking the questions, you can transmit a message without having to ask anything.

Here is a chart called a ‘decision tree’ for guessing a number between 0 and 7:



What are the yes/no decisions needed to ‘guess’ the number 5?

How many yes/no decisions do you need to make to work out any number?

Now look at something very fascinating. Underneath the numbers 0, 1, 2, 3... in the final row of the tree write the number in binary (see Activity 1).

Look closely at the tree. If no=0 and yes=1, what do you see?

In the number guessing game we try to choose questions so that the sequence of answers works out to represent the number in exactly this way.

Design your own decision tree for guessing numbers between 0 and 15.

Extra for experts: What kind of tree would you use to guess someone’s age? What about a tree to guess which letter is next in a sentence?

What's it all about?

A celebrated American mathematician (and juggler, and unicyclist) called Claude Shannon did a lot of experiments with this game. He measured the amount of information in bits—each yes/no answer is equivalent to a 1/0 bit. He found that the amount of "information" contained in a message depends on what you already know. Sometimes we can ask a question that eliminates the need to ask a lot of other questions. In this case the information content of the message is low. For example, the information in a single toss of a coin is normally one bit: heads or tails. But if the coin happens to be a biased one that turns up heads nine times out of ten, then the information is no longer one bit—believe it or not, it's less. How can you find out what a coin toss was with less than one yes/no question? Simple—just use questions like "are the next *two* coin tosses both heads?" For a sequence of tosses with the biased coin, the answer to this will be "yes" about 80%, of the time. On the 20% of occasions where the answer is "no," you will have to ask two further questions. But on average you will be asking less than one question per coin toss!

Shannon called the information content of a message "entropy". Entropy depends not only on the *number* of possible outcomes—in the case of a coin toss, two—but also on the *probability* of it happening. Improbable events, or surprising information, need a lot more questions to guess the message because they tell us more information we didn't already know—just like the situation of taking a helicopter to school.

The entropy of a message is very important to computer scientists. You cannot compress a message to occupy less space than its entropy, and the best compression systems are equivalent to a guessing game. Since a computer program is making the 'guesses', the list of questions can be reproduced later, so as long as the answers (bits) are stored, we can reconstruct the information! The best compression systems can reduce text files to about a quarter of their original size—a big saving on storage space!

The guessing method can also be used to build a computer interface that predicts what the user is going to type next! This can be very useful for physically disabled people who find it difficult to type. The computer suggests what it thinks they are likely to type next, and they just indicate what they want. A good system needs an average of only two yes/no answers per character, and can be of great assistance to someone who has difficulty making the fine movements needed to control a mouse or keyboard. This sort of system is also used in a different form to 'type' text on some cellphones.

Solutions and hints

The answer to a single yes/no question corresponds to exactly one bit of information—whether it is a simple question like “Is it more than 50?” or a more complex one like “Is it between 20 and 60?”

In the number-guessing game, if the questions are chosen in a certain way, the sequence of answers is just the binary representation of the number. Three is 011 in binary and is represented by the answers “No, yes, yes” in the decision tree, which is the same if we write no for 0 and yes for 1.

A tree you would use for someone’s age might be biased towards smaller numbers.

The decision about the letters in a sentence might depend upon what the previous letter was.