

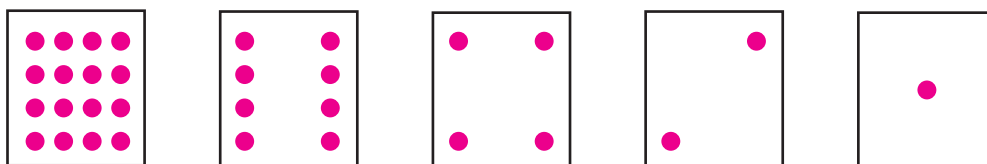
Représenter des informations

> D'APRÈS *COMPUTER SCIENCES UNPLUGGED* ([HTTP://CSUNPLUGGED.ORG](http://CSUNPLUGGED.ORG)), PROGRAMME D'ACTIVITÉS D'ÉVEIL POUR LES ÉLÈVES À PARTIR DE L'ÉCOLE PRIMAIRE, CRÉÉ PAR TIM BELL, IAN H. WITTEN ET MIKE FELLOWS, ADAPTÉ À L'UTILISATION EN CLASSE PAR ROBYN ADAMS ET JANE MCKENZIE, ILLUSTRÉ PAR MATT POWELL (NOUVELLE-ZÉLANDE). VERSION FRANÇAISE COORDONNÉE PAR L'ÉQUIPE D'INTERSTICES ([HTTP://INTERSTICES.INFO](http://INTERSTICES.INFO))

A Compter les points : l'écriture binaire des nombres

Les données de l'ordinateur sont stockées et transmises sous la forme d'une série de 0 et de 1. Comment peut-on représenter des mots et des nombres à l'aide de ces deux chiffres uniquement ?

Pour cette activité, il faut cinq cartes, avec des points sur le recto mais rien sur le verso, rangées dans l'ordre suivant.

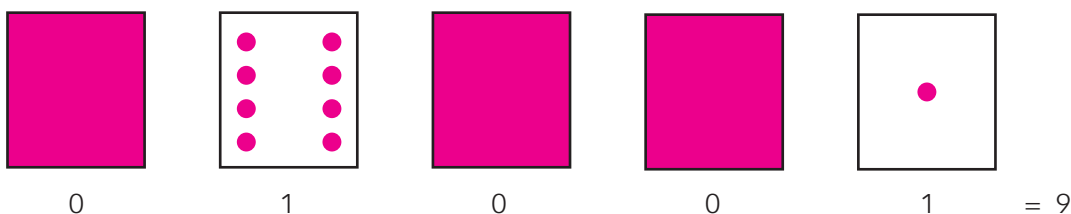


On commencera par faire remarquer aux élèves que chaque carte a deux fois plus de points que la carte située à sa droite. Ainsi, si l'on ajoute une carte à gauche, elle comportera 32 points.

On peut utiliser ces cartes pour représenter des nombres : il faut en retourner certaines et additionner les points qui restent visibles : 6 = cartes à 4 points et à 2 points ; 15 = cartes à 8, 4, 2 et 1 points ; 21 = cartes à 16, 4 et 1 points, etc.

Maintenant, on peut essayer de compter à partir de zéro. Lorsqu'une carte d'un nombre binaire n'est pas visible, elle est représentée par un 0. Lorsqu'elle est visible, elle est représentée par un 1. C'est le système de numération binaire : l'écriture des nombres dans ce système se fera uniquement avec les caractères 0 et 1, contrairement à l'écriture des nombres du système décimal qui utilise les chiffres de 0 à 9.

Ainsi, 01001 représente 9 dans le système décimal :



Sur ce principe, 17 s'écrit 10001 en binaire.

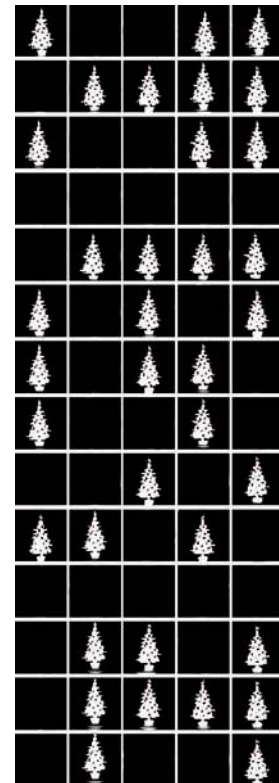
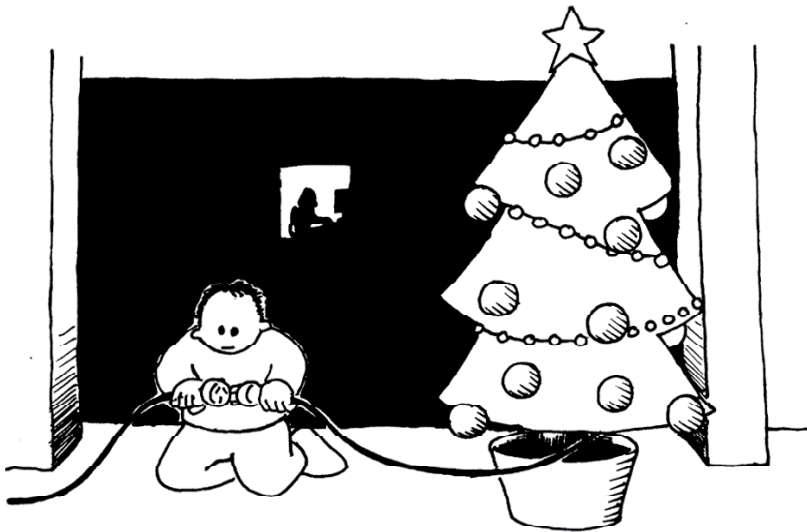
Exercice 1

- Place les 5 cartes de gauche à droite en commençant par celle à 16 points.
- Retourne les cartes de manière à laisser 5 points visibles, en conservant toujours les cartes dans le même ordre.
- Trouve comment obtenir 3, 12, 19.
- Y a-t-il plusieurs moyens d'obtenir un nombre ? (Non)
- Quel est le plus grand nombre que tu peux obtenir ? (31)
- Quel est le plus petit ? (0)
- Y a-t-il un nombre entre le plus grand et le plus petit que tu ne puisses pas obtenir ? (Non)

Voici donc une nouvelle méthode pour compter. Les ordinateurs utilisent uniquement le 0 et le 1. Tout ce que qu'on entend ou voit sur l'ordinateur – mots, images, nombres, films et même sons – est stocké à l'aide de ces deux chiffres uniquement. Grâce à ce système, on peut s'amuser à envoyer des messages secrets.

Exercice 2

Tom est pris au piège à l'étage supérieur d'un grand magasin. Noël approche et il veut rentrer à la maison avec ses cadeaux. Il a essayé d'appeler, et même de crier, mais il n'y a plus personne. Il peut voir de l'autre côté de la rue quelqu'un qui travaille sur son ordinateur tard ce soir. Comment pourrait-il attirer son attention ? Tom regarde autour de lui et il a alors une brillante idée : il peut utiliser les lumières de l'arbre de Noël pour lui envoyer un message ! Il trouve toutes les lumières et les branche de manière à pouvoir les allumer et les éteindre. Il utilise un code binaire simple, dont il est sûr que la personne de l'autre côté de la rue le comprendra. Pourras-tu le trouver ?



1	2	3	4	5	6	7	8	9	10	11	12	13
a	b	c	d	e	f	g	h	i	j	k	l	m
14	15	16	17	18	19	20	21	22	23	24	25	26
n	o	p	q	r	s	t	u	v	w	x	y	z

Les ordinateurs sont connectés à internet via un modem qui utilise aussi le système binaire pour envoyer des messages. La seule différence est que les modems téléphoniques utilisent des bips : un bip aigu pour le 1 et un bip grave pour le 0. Ces sons vont si vite que ce que nous entendons est un son continu et strident..

Comment faire maintenant pour compter au-delà de 31 ? Prenons la suite suivante : 1, 2, 4, 8, 16... En additionnant 1 + 2 + 4, on obtient 7. En additionnant 1 + 2 + 4 + 8 +16, on obtient 31. En comptant sur ses doigts, on ne pourra pas aller au-delà de 10. Par contre, en utilisant le système binaire et si chaque doigt représente une des cartes, on peut compter de 0 à 31, ce qui fait 32 nombres (le zéro aussi est un nombre !). Un doigt représente le 1 s'il est relevé ou le 0 s'il ne l'est pas. En utilisant nos deux mains, il est possible en fait de compter de 0 à 1023, c'est-à-dire 1024 nombres !

On observe une autre propriété intéressante des nombres binaires. Si nous travaillons en base dix (écriture décimale), placer un 0 à droite du nombre revient à le multiplier par 10. Par exemple, 9 devient 90, 30 devient 300. En ajoutant 0 à droite d'un nombre écrit en binaire, on le multiplie par deux : 1001 (9) -> 10010 (18).

Chacune des cartes que nous avons utilisées jusqu'à présent représente un « bit » sur l'ordinateur (« bit » est la contraction de *binary digit*, qui signifie « chiffre binaire»). Ainsi, le code alphabétique que nous avons utilisé précédemment peut être représenté grâce à seulement cinq cartes, ou bits. Cependant, un ordinateur doit savoir si les lettres sont en majuscules ou non et doit également reconnaître les chiffres, la ponctuation et les caractères accentués ou spéciaux comme \$ ou &.

L'ordinateur a besoin de 7 bits pour stocker les 128 caractères d'un clavier. La plupart des ordinateurs utilisent aujourd'hui le système de représentation ASCII (*american standard code for information interchange*, qui signifie « code américain normalisé pour l'échange d'informations») qui est basé sur l'utilisation de ce nombre de bits pour chaque caractère. Mais certains pays non anglophones utilisent des codes plus longs.

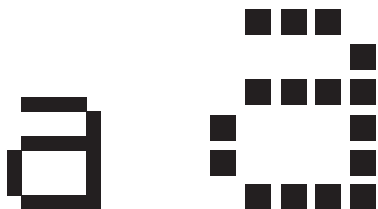
À retenir. Aujourd'hui, les ordinateurs utilisent le système binaire pour représenter les informations. Ce système tient son nom du fait qu'il n'utilise que deux chiffres : 0 et 1. Il est également connu sous le nom « base deux ». Chaque 0 ou 1 est appelé un « bit » (*binary digit*). Un bit est généralement représenté dans la mémoire principale de l'ordinateur par un transistor qui est activé ou désactivé, ou un condensateur qui est chargé ou déchargé. Lorsque des données doivent être transmises par une ligne téléphonique ou par liaison radio, des sons graves et aigus représentent les 1 et les 0. Sur des disques magnétiques et des bandes, les bits sont représentés par la direction (nord-sud ou sud-nord) d'un champ magnétique sur une surface. Sur les CD audio, CD-Rom et DVD, les bits sont stockés de manière optique : la partie de la surface correspondant au bit reflète ou ne reflète pas la lumière. Un bit ne peut pas représenter beaucoup d'informations à lui tout seul. Ils sont donc regroupés, généralement par 8, ce qui représente des nombres allant de 0 à 255. Un groupe de 8 bits s'appelle un octet. La vitesse d'un ordinateur dépend du nombre de bits qu'il peut traiter simultanément. Par exemple, un ordinateur 32 bits peut traiter des nombres de 32 bits en une seule opération, alors qu'un ordinateur 16 bits doit traiter les nombres de 32 bits en plusieurs fois, ce qui le ralentit. Enfin, les bits et les octets sont les seuls éléments dont l'ordinateur a besoin pour stocker et transmettre des nombres, du texte et tout autre type d'informations.

Dans les activités suivantes, nous verrons comment d'autres types d'informations peuvent être représentés sur un ordinateur.

B La couleur par les nombres : représentation d'une image

À quoi servent les fax (télécopieurs) ? Dans quelles circonstances les ordinateurs ont-ils besoin de stocker des images (un programme de dessin, un jeu avec des graphiques, ou un système multimédia) ? Comment les ordinateurs peuvent-ils stocker des images alors qu'ils n'utilisent que des nombres ?

Les écrans d'ordinateur sont divisés en une grille de petits points appelés « pixels » (*picture elements*, qui signifient « éléments d'image »). Dans une image en noir et blanc, chaque pixel est soit noir soit blanc. La lettre *a* a été agrandie ci-dessous pour que l'on voie bien les pixels.



Lorsqu'il stocke une image, l'ordinateur enregistre l'emplacement des points noirs et des points blancs. L'image ci-dessous nous montre comment une image peut être représentée par des nombres.

	■	■	■		1, 3, 1
				■	4, 1
	■	■	■	■	1, 4
■				■	0, 1, 3, 1
■				■	0, 1, 3, 1
	■	■	■	■	1, 4

La première ligne contient un pixel blanc, trois noirs puis un blanc. Ainsi, la première ligne est représentée par 1, 3, 1. Le premier nombre représente toujours le nombre de pixels blancs. Si le premier pixel est noir, la ligne commencera par un 0.

À retenir. Un télécopieur n'est en fait qu'un simple ordinateur qui scanne une page en noir et blanc en 1 000 X 2 000 pixels. Ces pixels sont envoyés grâce à un modem vers un autre télécopieur, qui les imprime sur une page. Souvent, les images télécopiées ont de grands blocs de pixels blancs (les marges par exemple) ou noirs (lignes horizontales par exemple). Les images en couleur comportent également de nombreuses répétitions. Pour économiser suffisamment d'espace pour stocker ces images, les programmeurs peuvent utiliser différentes techniques de compression. La méthode utilisée dans cette activité, appelée « codage par plages », est efficace pour compresser les images. Si les images n'étaient pas compressées, il faudrait beaucoup plus de temps pour les transmettre et un espace plus important pour les stocker. Par exemple, les images de fax sont généralement compressées à environ un septième de leur taille initiale. Sans compression, leur transmission prendrait sept fois plus de temps ! Les photographies et autres images sont souvent compressées à un dixième, et même à un centième, de leur taille initiale (grâce à une autre technique).

C Peux-tu répéter ? Compression de texte

Les ordinateurs stockent et transmettent de grandes quantités de données. Pour éviter d'utiliser trop d'espace de stockage ou de mettre trop de temps à envoyer les informations par le modem, ils compressent le texte d'une manière similaire à celle présentée ci-dessous.

Prenons les paroles de cette célèbre chanson :

Vive le vent, vive le vent
Vive le vent d'hiver
Qui s'en va en sifflant
Dans les grands sapins verts

On commence par chercher les séquences de deux lettres ou plus qui se répètent dans ce poème, puis on remplace ces groupes de lettres répétés par des cases, comme ci-dessous :

sifflant soufflant
↓
sifflant sou

Exercice 1

Il manque de nombreuses lettres et de nombreux mots dans ce poème. Peux-tu remplir les blancs pour le compléter correctement ? Tu les trouveras dans les cases indiquées par la pointe de la flèche.

Un peti cha gr s

qui mangea du z

sur un tapis

sa lui d

ce n'est pas o

de r

Exercice 2

Imagine que tu es un ordinateur et que tu dois stocker autant d'informations que possible sur ton disque. Raye tous les groupes d'au moins deux lettres qui se répètent. Ces groupes ne sont plus nécessaires puisqu'on peut les remplacer par des flèches. Le but est d'obtenir le plus grand nombre possible de lettres rayées.

L'escargot
Un escargot s'en allait à la foire
Pour s'acheter une paire de souliers,
Quand il arriva, il faisait déjà nuit noire
Il s'en retourna nu-pieds.
Un escargot s'en allait à l'école
Car il voulait apprendre à chanter
Quand il arriva, ne vit que des herbes folles
C'était les vacances d'été.
Un escargot s'en allait en vacances
Pour visiter l'Inde et le Japon
Au bout de sept ans, il était toujours en France
Entre Dijon et... Lyon!

À retenir. La capacité de stockage des ordinateurs augmente à un rythme incroyable (environ 1 million de fois sur un ordinateur classique au cours des vingt-cinq dernières années), mais nous trouvons toujours plus à y stocker (livres entiers et même des bibliothèques, musique, vidéos...). Des fichiers volumineux posent également un problème pour leur transmission par internet car ils sont plus longs à télécharger. On essaie aussi de construire des ordinateurs toujours plus petits : il est même possible de stocker beaucoup d'informations sur des téléphones mobiles. Cependant, il existe des solutions à ces problèmes. Plutôt que d'acheter plus d'espace de stockage ou un modem plus rapide, nous pouvons compresser les données de manière qu'elles occupent moins d'espace. L'ordinateur effectue en général automatiquement cette compression et cette décompression des données.

De nombreuses méthodes de compression ont été imaginées. Celle utilisée pour cette activité, basée sur le principe des flèches dirigées vers des occurrences de portions de texte, est souvent appelée « algorithme de Ziv-Lempel » ou « algorithme LZ », car il a été découvert par deux professeurs israéliens dans les années 1970. Cet algorithme s'applique à tout langage et peut facilement diminuer de moitié la taille des données compressées. Il est parfois appelé « zip » sur les PC et sert aussi pour les images GIF et les modems haut débit. Selon d'autres méthodes, les lettres utilisées le plus souvent sont représentées par un code plus court que les autres. C'est le principe du morse.

D Tour de cartes : détection et correction des erreurs

Lorsque les données sont stockées sur un disque ou transmises d'un ordinateur à un autre, il arrive parfois que les données soient modifiées accidentellement. Cette activité propose une méthode pour savoir si les données ont été altérées et pour pouvoir les corriger.

Nous avons besoin d'un ensemble de cartes identiques, à deux faces. Pour la présentation, il vaut mieux utiliser des cartes magnétiques plates qui ont une couleur différente sur chaque face (des aimants pour réfrigérateur par exemple).

1. Demander à un élève de disposer les cartes en carré de 5 X 5, avec une répartition aléatoire des faces visibles. Ajouter soi-même une autre colonne et une autre ligne, « pour compliquer un peu les choses ». Il faut choisir les cartes supplémentaires de manière à obtenir un nombre pair de cartes colorées sur chaque ligne et chaque colonne.
2. Fermer les yeux et demander à un autre élève de retourner une seule carte. La ligne et la colonne de la carte contiennent alors un nombre impair de cartes colorées et l'on peut ainsi identifier la carte retournée.

Expliquer l'astuce

1. En travaillant en binôme, les élèves disposent leurs cartes en carré de 5 X 5.
2. Combien de cartes colorées y a-t-il sur chaque ligne et chaque colonne ? S'agit-il d'un nombre pair ou impair ? Souvenez-vous que 0 est un nombre pair.

3. Ajouter maintenant une sixième carte à chaque ligne, en s'assurant que le nombre de cartes colorées reste pair. Cette carte supplémentaire est appelée la « carte de la parité », qui correspond à ce que les informaticiens appellent le « bit de parité ».

4. Ajouter une sixième ligne de cartes en bas, afin d'obtenir un nombre pair de cartes colorées dans chaque colonne.

5. Maintenant, retourner une carte. Que remarque-t-on sur la ligne et la colonne de la carte retournée ? (Elles comportent un nombre impair de cartes colorées.) La carte de la parité illustre une erreur.

Un exemple tirée de la vie réelle. Des techniques de vérification similaires sont utilisées pour les codes d'identification des livres. Les ouvrages publiés comportent un code à dix chiffres que l'on trouve généralement sur la quatrième de couverture. Le dernier chiffre est un chiffre (ou somme) de contrôle, tout comme les bits de parité dans l'exercice précédent. Cela signifie que si l'on commande un livre d'après son numéro ISBN (*international standard book number*, qui signifie « numéro international normalisé des livres »), l'éditeur peut vérifier qu'il n'y a pas eu d'erreur. Il regarde simplement la somme de contrôle.

Voici comment obtenir la somme de contrôle :

– Multiplier le premier chiffre par dix, le deuxième par neuf, le troisième par huit, etc. jusqu'au neuvième chiffre multiplié par deux. Additionner tous les résultats obtenus. Par exemple, le numéro ISBN 0-13-911991-4 donne le résultat suivant : $(0 \times 10) + (1 \times 9) + (3 \times 8) + (9 \times 7) + (1 \times 6) + (1 \times 5) + (9 \times 4) + (9 \times 3) + (1 \times 2) = 172$

– Diviser ensuite ce résultat par onze. $172 \div 11 = 15$, reste 7

– Si le reste est égal à zéro, alors la somme de contrôle est zéro ; sinon, soustraire le reste à 11 pour obtenir le chiffre de contrôle : $11 - 7 = 4$. S'agit-il bien du dernier chiffre du numéro ISBN ? Oui ! Si ce n'était pas le cas, alors nous saurions qu'une erreur s'est produite.

À retenir. Un ordinateur qui reçoit une instruction doit vérifier que les données qui lui arrivent n'ont pas été altérées par une interférence électrique sur la ligne. Parfois, lorsqu'une erreur a été transmise, les données originales peuvent être renvoyées mais ce n'est pas toujours le cas, par exemple si le disque ou la bande ont été physiquement endommagés, exposés à un rayonnement magnétique ou électrique ou à la chaleur. Si les données sont envoyées par une sonde spatiale lointaine, l'attente d'une retransmission en cas d'erreur serait interminable ! Nous devons pouvoir reconnaître si les données ont été altérées (détecter l'erreur) et reconstituer les données originales (corriger l'erreur).

Les ordinateurs utilisent la même technique que celle utilisée pour le tour de cartes. En plaçant les bits en lignes et colonnes imaginaires et en ajoutant les bits de parité à chaque ligne et colonne, nous pouvons non seulement détecter une erreur, mais aussi la localiser. Le bit défectueux est réparé : l'erreur est corrigée. Bien sûr, les ordinateurs utilisent souvent des systèmes de contrôle des erreurs plus complexes, capables de détecter et de corriger des erreurs multiples. Le disque dur d'un ordinateur réserve une grande partie de son espace à la rectification des erreurs, de manière à pouvoir continuer à travailler même si certaines parties du disque sont défectueuses.

E Vingt devinettes : théorie de l'information

Comment peut-on mesurer la quantité d'informations contenues dans un livre ? Le nombre de pages ou de mots est-il important ? Un livre contient-il plus d'informations qu'un autre ?

Les informaticiens mesurent les informations d'après le degré de nouveauté du message (ou du livre !). Si l'on entend quelque chose que l'on connaît déjà, par exemple, un ami qui se rend toujours à l'école à pied dit « Je suis venu à l'école à pied aujourd'hui », le message ne comporte rien de nouveau. Si cet ami avait dit « Aujourd'hui, je suis venu en hélicoptère », alors oui, ça aurait été nouveau, et nous aurions eu dans ce message beaucoup d'informations.

Comment mesurer le degré de nouveauté d'un message ? L'une des méthodes consiste à observer la difficulté avec laquelle on devine l'information. Si cet ami dit « Devine comment je suis venu à l'école aujourd'hui ? » et qu'il est venu à pied, il sera très facile de trouver la bonne réponse. Mais il faudra plusieurs essais avant de deviner que c'est en hélicoptère, et encore plus si c'est en navette spatiale. La quantité d'informations contenue dans un message se mesure d'après la facilité ou la difficulté à les deviner. Le jeu suivant donne une idée de la complexité à deviner l'information.

Voici une adaptation du jeu de vingt questions. Un élève vient au tableau. Les autres doivent lui poser des questions auxquelles il ne peut répondre que par oui ou par non, jusqu'à ce qu'ils trouvent la bonne réponse.

Exemple : je pense à

un nombre compris entre 1 et 100,

un nombre compris entre 1 et 1 000,

un nombre compris entre 1 et 1 000 000,

un nombre quelconque,

une suite de six nombres qui respecte une séquence (adaptée au groupe).

Comptez le nombre de questions posées. C'est la mesure de la valeur des « informations ».

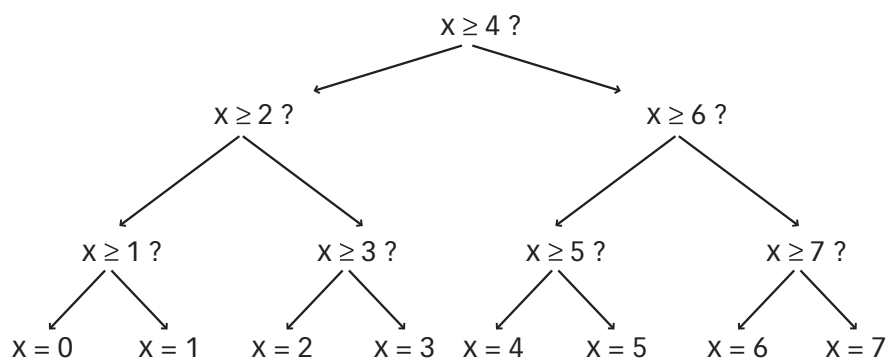
Quelles stratégies avez-vous appliquées ? Quelles étaient les plus efficaces ? Il ne faut que sept propositions pour deviner un nombre compris entre 1 et 100 si vous divisez l'intervalle par deux à chaque question. Par exemple :

Le nombre est-il inférieur ou égal à 50 ? Oui.
 Le nombre est-il inférieur ou égal à 25 ? Non.
 Le nombre est-il inférieur ou égal à 37 ? Non.
 Le nombre est-il inférieur ou égal à 43 ? Oui.
 Le nombre est-il inférieur ou égal à 40 ? Non.
 Le nombre est-il inférieur ou égal à 41 ? Non.
 Ce ne peut être que 42 ! Oui !

On remarque que si on étend l'intervalle jusqu'à 1000, il ne faut pas 10 fois plus de propositions, mais 3 seulement. Chaque fois que l'intervalle double, une seule question supplémentaire suffit à trouver la réponse. À la suite de cet exercice, il serait intéressant de faire jouer les enfants au jeu du Mastermind.

Exercice : l'arbre de décision

Si vous connaissez déjà la stratégie pour poser les questions, vous pouvez transmettre un message sans poser une seule question. Voici un schéma appelé « arbre de décision » pour deviner un nombre entre 0 et 7.



- Quelles décisions oui/non sont nécessaires pour « deviner » le nombre 5 ?
- Combien de décisions oui/non sont nécessaires pour trouver un nombre quelconque ?
- Observe à présent ce qui est fascinant : sous les nombres 0, 1, 2, 3... du dernier niveau de l'arbre, écris les nombres en binaire (voir **Activité A**).
- Regarde l'arbre attentivement. Si non = 0 et oui = 1, que constates-tu ?
- Dans le jeu de devinette des nombres, nous essayons de choisir les questions pour que la suite des réponses permette de représenter le nombre de cette façon.
- Construis ton propre arbre de décision pour deviner les nombres compris entre 0 et 15.

À retenir. Claude Shannon, célèbre mathématicien américain, a fait de nombreuses expériences avec ce jeu. Il a mesuré la quantité d'informations en bits (chaque oui/non étant équivalent à un bit 1/0). Il a trouvé que la quantité d'« informations » contenues dans un message dépend de ce que l'on sait déjà. Parfois, nous pouvons poser une question qui élimine le besoin d'en poser de nombreuses autres. Dans ce cas, le contenu informatif du message est restreint. Par exemple, lancer une pièce représente normalement une information d'un bit : pile ou face. Mais s'il se trouve que la pièce est faussée et qu'elle tombe neuf fois sur dix sur face, alors l'information n'est plus un bit mais moins, aussi incroyable que cela paraisse. Comment connaître le résultat d'un lancer en posant moins d'une question oui/non ? C'est simple : avec une seule question comme « Les deux prochains lancers tomberont-ils sur face ? », pour une suite de lancers avec une pièce faussée, la réponse serait « oui » à 80%. Dans les 20% des cas restant, lorsque la réponse est « non », vous devrez poser deux autres questions. Mais, en moyenne, vous aurez posé moins d'une question par lancer de pièce ! Claude Shannon appelait ce contenu informatif d'un message « l'entropie ». L'entropie dépend non seulement du nombre de résultats possibles (dans le cas du lancer de pièce : deux), mais également de la probabilité de l'événement. Les événements improbables, ou informations nouvelles, nécessitent beaucoup plus de questions pour deviner le message car ils transmettent plus d'informations encore inconnues.

L'entropie d'un message est très importante pour les informaticiens. Les meilleurs systèmes de compression peuvent réduire les fichiers texte à un quart environ de leur taille initiale (une grande économie sur l'espace de stockage !)

La méthode des devinettes peut également servir à construire une interface informatique qui prédit quel texte l'utilisateur s'apprête à saisir ! Cela peut être très utile pour les personnes présentant un handicap physique qui ont des difficultés à taper. L'ordinateur suggère ce qu'il pense que l'utilisateur va taper ensuite et ce dernier n'a qu'à indiquer si c'est bien ce qu'il souhaite. Un bon système nécessite en moyenne seulement deux questions oui/non par caractère et peut être d'une grande utilité pour quelqu'un qui peine à réaliser les petits mouvements précis nécessaires pour manipuler la souris ou le clavier. Ce type de système est également utilisé sous différentes formes pour « saisir » du texte sur les téléphones mobiles.