

# Entropie, information : un concept protéiforme

par Roger Balian

Le mot d'*entropie* a été forgé en 1865 par le physicien allemand Clausius à partir de la racine grecque *τροπή* qui évoque l'idée de transformation ou de retour en arrière. Il introduisit cette grandeur, comme nous le préciserons plus loin, afin de caractériser mathématiquement l'irréversibilité de processus physiques tels qu'une transformation de travail en chaleur. Depuis cette époque, tel le dieu grec Protée, le concept d'entropie n'a cessé de se métamorphoser, d'échapper à qui voulait le saisir, et comme lui il s'est référé à la prévision. Son importance s'est affirmée, et son champ d'application s'est étendu progressivement à la chimie, aux mathématiques, à l'informatique, à la biologie, à l'économie. Nous verrons en particulier que la notion de *quantité d'information*, utile en théorie de la communication ou en informatique, est étroitement apparentée à celle d'entropie, exemple remarquable d'unification entre champs du savoir fort éloignés en apparence.

En fait, il a fallu près de deux siècles pour réaliser cette unification, tant le cheminement de la science a été tortueux dans le domaine qui nous concerne ici. Ainsi, ce que nous appelons le "premier" principe de la thermodynamique, la conservation de l'énergie, n'a été reconnu qu'une vingtaine d'années après l'énoncé, en 1824, du "deuxième" principe par Sadi Carnot. C'est plus tôt encore, en 1811, que la loi de l'écoulement de la chaleur a été formulée par Joseph Fourier, alors que la nature physique de la chaleur n'a été élucidée qu'à la fin du XIX<sup>ème</sup> siècle. Le concept de quantité d'information a été inspiré par celui d'entropie, plus ancien d'un siècle ; c'est pourtant lui qui apparaît aujourd'hui comme le plus fondamental, et c'est par lui que nous commencerons.

## Quantité d'information ; manque d'information

La théorie de la communication a pour objet d'optimiser la transmission à distance de messages, qu'ils soient écrits, sonores ou visuels. Elle a été fondée en 1948 par Claude Shannon et Warren Weaver, mathématiciens américains qui travaillaient alors à l'amélioration des systèmes de téléphonie. Il leur fallait dans ce but comparer les performances de divers processus de transfert d'information, et une telle comparaison nécessitait de chiffrer la quantité d'information véhiculée par les messages à transmettre. Un problème semblable se pose en informatique lorsqu'on souhaite stocker dans une bande magnétique, un disque ou une mémoire d'ordinateur divers documents, qu'il s'agisse de données numériques, de textes, de musique, d'images réelles ou virtuelles. Il importe alors d'occuper le moins de place possible dans la mémoire, et on ne peut y parvenir systématiquement que si l'on sait évaluer numériquement la quantité d'information contenue dans ces documents.

On remarque d'abord que le message à transmettre (ou le document à stocker) n'a pas de signification en soi, mais qu'il est extrait d'une famille de messages, tous a priori susceptibles d'être émis. Ainsi, le nom du vainqueur d'une compétition d'athlétisme est extrait de la liste des concurrents. Chacun de ceux-ci, désigné par  $m$ , a une certaine probabilité  $p_m$  de vaincre. En apprenant que c'est l'athlète  $m$  qui a gagné, on acquiert une certaine quantité d'information  $I_m$ . Celle-ci doit être d'autant plus grande que  $m$  avait une probabilité  $p_m$  plus faible de gagner :

si  $m$  était un outsider, sa victoire inattendue constitue une nouvelle importante ; si au contraire  $m$  était le grand favori, sa victoire ne nous apporte guère d'information. La quantité d'information  $I_m$  associée à un message  $m$  est donc une fonction décroissante de sa probabilité  $p_m$ . De plus, cette fonction doit être additive : si un message est constitué de deux parties statistiquement indépendantes (comme les résultats de deux épreuves), la quantité d'information qu'il apporte doit être la somme des quantités d'information apportées par chacune de ses deux parties, tandis que sa probabilité est le produit des probabilités de ses parties. Il en résulte que la quantité d'information véhiculée par un message  $m$  de probabilité  $p_m$  vaut

$$I_m = \log_2 \frac{1}{p_m}. \quad (1)$$

Le choix de la base 2 pour le logarithme définit l'unité d'information, le *bit* : si le message se réduit à l'annonce du résultat d'un tirage à pile ou face, sa probabilité vaut  $p_m = 1/2$ , et il transporte un bit d'information.

Plaçons-nous avant transmission (ou avant mise en mémoire). Le destinataire ne connaît encore que l'ensemble des messages  $m$  susceptibles d'être reçus. Son incertitude  $I$ , sa perplexité, est mesurée par la valeur moyenne de la quantité d'information qu'il va acquérir lors de la réception de l'un d'eux. Ce *manque d'information*  $I$  s'obtient donc en pondérant l'information (1) associée à chaque message  $m$  par la probabilité  $p_m$  de recevoir ce message. Dans le cas particulier où les messages, en nombre total  $W$ , ont tous même probabilité  $p_m = 1/W$ , on obtient

$$I = \sum_m p_m I_m = \log_2 W. \quad (2)$$

Ce concept de manque d'information est utilisé en pratique pour optimiser le transfert de messages ou le stockage de documents. Dans les deux cas, il est nécessaire de coder le message ou le document par des signaux, par exemple électriques ou magnétiques, qui seront émis ou mémorisés. L'emploi de l'expression (2) a permis à Shannon et Weaver de démontrer une propriété remarquable : il existe pour toute famille de messages un mode de *codage optimal* qui permet de faire transiter un maximum d'informations dans une ligne de communication donnée, malgré des pertes éventuelles de signaux, ou de rendre minimal l'espace occupé par des documents dans une mémoire. La télévision "numérique" en fournit un exemple. Le codage optimal dépend des messages à transmettre : ce n'est pas le même pour la parole (téléphone) ou pour la musique (radio) ; ce n'est pas le même pour des textes écrits (fax) ou pour des photos.

Les notions de mémorisation, de communication et de codage sont devenues essentielles en *biologie* depuis que l'on a découvert le rôle de l'information génétique dans les mécanismes moléculaires du vivant. La séquence des nucléotides qui se succèdent le long des chaînes d'ADN forme un message, mémorisé dans les chromosomes ; ce message est codé par un alphabet de 4 lettres, les 4 nucléotides. Par ailleurs, les protéines assurent les principales fonctions vitales grâce à leur structure, elle-même gouvernée par la séquence d'acides aminés qui constitue chaque protéine. Ce type de séquence forme encore un message, codé cette fois par un alphabet de 20 lettres, les 20 acides aminés. C'est la transcription du message de l'ADN qui engendre l'enchaînement des acides aminés lors de la synthèse d'une protéine dans la cellule.

## Probabilités et incertitude

La formule (2) permet d'associer à toute loi de probabilité un nombre  $I$ . Dans ce contexte (fig.1, deuxième colonne), les probabilités  $p_m$  sont affectées à des événements aléatoires quelconques  $m$ , et le nombre  $I$  fournit une mesure naturelle de

l'étalement de la loi de probabilité, ou une *mesure de l'incertitude* associée. Pour un nombre donné d'événements,  $I$  est nul en cas de certitude, lorsque l'une des probabilités  $p_m$  vaut 1, les autres étant nulles ;  $I$  est maximal, et prend la valeur (2), lorsque les  $W$  événements ont la même probabilité. Cette valeur (2) croît avec  $W$  : l'incertitude sur un événement à venir est d'autant plus grande qu'il y a plus d'événements possibles, avec des probabilités  $1/W$  égales.

La théorie de l'information rend ainsi quantitative la notion d'incertitude liée à une description probabiliste, en associant le nombre  $I$  au caractère incomplet de notre connaissance des événements  $m$  susceptibles de se produire. L'existence de cette mesure de l'incertitude a donné naissance à un critère dit d'entropie maximale, utilisé pour affecter des probabilités à des événements aléatoires, par exemple pour interpréter des images imparfaites : parmi les diverses lois de probabilité possibles, on choisit celle qui rend l'incertitude  $I$  maximale afin d'éviter des affirmations infondées.

	<b>Communications</b>	<b>Probabilités</b>	<b>Physique statistique</b>
$m$	message	événement	configuration
$\{p_m\}$	ensemble des messages	loi de probabilité	état thermodynamique
$I, S$	manque d'information	incertitude, étalement	entropie, désordre

Fig. 1. Correspondance entre théorie de la communication, probabilités et physique statistique

## L'entropie en thermodynamique

Venons-en au concept d'entropie, qui paraît à première vue étranger à celui de quantité d'information quoique tous deux soient issus de considérations technologiques. L'information a été introduite pour améliorer le rendement des systèmes de communication. La thermodynamique, elle, dont l'entropie est l'une des grandeurs fondamentales, est née en 1824 des "Réflexions sur la puissance motrice du feu" de Carnot, en liaison avec le rendement des machines à vapeur. Elle est devenue ensuite la *science des transformations* : elle fournit les lois générales qui régissent toute transformation susceptible de survenir dans la matière à l'échelle macroscopique, c'est-à-dire à notre échelle. Ces transformations peuvent être associées à des échanges de chaleur, à des phénomènes mécaniques, électromagnétiques ou encore chimiques. Certaines d'entre elles sont *réversibles*. Par exemple, un trampoline dont on néglige l'amortissement retourne périodiquement à son état initial ; les échanges alternent entre l'énergie élastique des ressorts, l'énergie cinétique du sauteur (qui croît avec sa vitesse verticale) et son énergie potentielle de gravité (qui croît avec son altitude). Il existe des moteurs électriques pouvant fonctionner en dynamo ; les énergies mécanique et électrique s'y transforment réversiblement l'une en l'autre. D'autres transformations sont *irréversibles*, comme la transformation d'énergie mécanique en chaleur dans un frein de voiture, comme un flux spontané de chaleur qui ne peut se s'écouler que dans un sens, du chaud vers le froid, ou comme une combustion.

La thermodynamique repose sur deux grands principes. Le premier principe est basé sur le concept d'énergie : dans toute transformation d'un système isolé, n'échangeant rien avec l'extérieur, l'énergie est *conservée*. Le deuxième principe concerne l'*irréversibilité* des transformations, et fait intervenir une grandeur qui ne se conserve pas, l'entropie. Celle-ci, notée  $S$ , est une fonction des variables qui caractérisent à l'échelle macroscopique l'état du système considéré. Pour un fluide

pur homogène, ces variables sont par exemple le volume, la masse et la température ; pour un système constitué de plusieurs parties,  $S$  s'obtient en ajoutant les entropies de chacune de ces parties. Le deuxième principe exprime qu'aucune transformation ne peut faire décroître  $S$  dans un système isolé.

Ce principe affirme l'existence de l'entropie mais ne fournit pas sa valeur en fonction des autres variables macroscopiques. Sa détermination expérimentale utilise des transformations réversibles où l'entropie est conservée.

La notion d'entropie a d'abord été utilisée pour caractériser mathématiquement l'irréversibilité de phénomènes thermiques. Considérons un processus où une source chaude cède de la chaleur, dont une partie est transformée en travail et dont le reste s'écoule vers une source froide. A la conservation de l'énergie s'ajoute la contrainte exprimant la croissance de l'entropie totale : le corps froid doit gagner plus d'entropie que le corps chaud n'en perd. Cette inégalité fournit une limite supérieure au travail que l'on peut extraire du processus, donc au rendement des machines thermiques. On a reconnu dans la seconde moitié du XIX<sup>ème</sup> siècle que la croissance de l'entropie  $S$  d'un système isolé mesure plus généralement le degré d'irréversibilité des processus. Ainsi, l'entropie d'une tasse de café au lait est supérieure à la somme des entropies du café et du lait séparés ; la différence est l'entropie de mélange. De même, en chimie, une combustion fait croître l'entropie.

## L'entropie en physique statistique

L'existence et les propriétés de l'entropie postulées par le deuxième principe en accord avec l'expérience demeurent quelque peu mystérieuses. Que signifie cette grandeur, dépendant des variables d'état de la matière, dont la croissance est associée à l'irréversibilité, c'est-à-dire à la "*flèche du temps*" ? Le concept d'énergie, abstrait, peut devenir intuitif grâce à sa propriété de conservation, mais celui d'entropie, plus abstrait encore, s'appréhende difficilement. Cette question n'a cessé de hanter Ludwig Boltzmann (1844-1906) qui a contribué à clarifier la nature de l'entropie et à en interpréter les variations.

Ses travaux reposent sur une idée-force, de type *réductionniste* : il faut chercher à expliquer les propriétés macroscopiques des objets en s'appuyant sur leur structure microscopique. Nous savons aujourd'hui qu'à l'échelle de l'angström (un dix-millième de micron) la matière qui nous entoure est constituée de molécules, d'atomes ou d'ions. Les lois, relativement simples, qui régissent ces particules sont *les mêmes* pour tous les matériaux, et l'immense variété des propriétés observées à notre échelle macroscopique résulte d'agencements divers entre les constituants élémentaires. Le *nombre immense* de ces derniers implique l'emploi de méthodes d'étude statistiques ; par exemple, une sphère d'air de 1 micron de diamètre contient 100 millions de molécules. Une science nouvelle, la physique statistique (ou mécanique statistique), s'est développée sur ces bases depuis le milieu du XIX<sup>ème</sup> siècle avec un succès considérable.

La physique statistique est d'abord parvenue à interpréter à l'échelle microscopique les *concepts de la thermodynamique* et à en *expliquer les principes* généraux. Ainsi, la pression d'un gaz s'identifie à une force moyenne exercée par les molécules en raison de leur mouvement et de leurs chocs sur les parois. La température s'interprète comme la manifestation macroscopique de l'agitation désordonnée des constituants microscopiques, et la chaleur comme l'énergie cinétique associée. Le premier principe résulte alors de la conservation à l'échelle microscopique de l'énergie mécanique. Un second but de la physique statistique consiste à déduire les *lois du comportement* des divers matériaux de leur structure microscopique. Ainsi, pour les gaz, l'équation d'état empirique de Mariotte-Boyle, les équations dynamiques de Navier-Stokes, découlent des lois simples du mouvement des molécules.

Le passage d'une échelle à l'autre entraîne une modification qualitative radicale des concepts et des propriétés. La microphysique est discontinue, probabiliste, linéaire, réversible, alors que la physique macroscopique qui en découle est continue, déterministe, non linéaire, irréversible. C'est la compréhension de la remarquable *émergence* de tels comportements nouveaux qui rend fructueux le réductionnisme.

Afin d'appréhender ainsi les fondements microscopiques du deuxième principe et la signification de l'entropie, notons d'abord qu'un état thermodynamique, bien que caractérisé par la donnée de quelques variables macroscopiques, est à l'échelle microscopique mal connu. Par exemple, la thermodynamique d'un gaz repose sur la spécification de trois variables, le volume, la masse et la température, qui à l'échelle microscopique déterminent seulement le volume offert aux molécules, leur nombre et leur énergie cinétique totale. Il existe une quantité gigantesque, notée  $W$ , de *configurations microscopiques* compatibles avec les variables thermodynamiques : elles diffèrent les unes des autres par le détail des positions et vitesses des molécules. Ces dernières variables microscopiques, en nombre bien plus immense encore que celui des molécules, sont inaccessibles. Puisque nous ne savons pas dans quelle configuration particulière se trouve un système dont les variables thermodynamiques sont données, il est naturel de le décrire par une approche statistique où l'on affecte à chacune des  $W$  configurations possibles la même probabilité  $1/W$ .

La physique statistique a ainsi la structure d'une théorie probabiliste : chaque configuration correspond à un événement, un état thermodynamique à une loi de probabilité (fig.1, dernière colonne). Bien qu'un tel état paraisse bien défini à notre échelle, il fait intervenir à l'échelle microscopique une incertitude  $I$ , mesurée en bits par la formule (2). On peut démontrer que les propriétés de cette incertitude  $I$  sont les mêmes que toutes celles que la thermodynamique a empiriquement attribuées à l'entropie  $S$ . Par suite, les deux grandeurs  $I$  et  $S$  s'identifient, à un facteur près :

$$S = k \log_2 W, \quad k = 0,96.10^{-23} \text{ JK}^{-1} \text{ bit}^{-1}. \quad (3)$$

Le nombre  $k$ , dit constante de Boltzmann, est déterminé par le choix de l'unité d'entropie, le joule par kelvin ( $\text{JK}^{-1}$ ), unité usuelle adaptée à notre échelle. Sa petitesse va de pair avec l'immensité des valeurs de  $W$ , qui vaut  $10^{3.10^{22}}$  lorsque  $S = 1 \text{ JK}^{-1}$ . Puisque  $1/k$  est de l'ordre du nombre d'Avogadro  $6.10^{23} \text{ mol}^{-1}$ , l'incertitude  $I = S/k$  sur un système macroscopique est de l'ordre de grandeur du nombre de ses constituants élémentaires.

La formule (3) a été gravée sur la tombe de Boltzmann à Vienne. Il l'avait imaginée dès 1877 à la suite d'efforts obstinés et solitaires, à une époque où la plupart des savants se défiaient de l'atomisme en l'absence de preuve expérimentale de la structure discontinue de la matière. Elle est remarquable à plus d'un titre. Elle clarifie la nature de l'entropie, grandeur dont la thermodynamique se contentait de postuler empiriquement l'existence. Elle rend accessible à la théorie, pour des systèmes suffisamment simples, le calcul de leur entropie en le ramenant à un dénombrement de configurations ; elle ouvre ainsi la voie à la prévision de propriétés physiques spécifiques des matériaux à partir de leur structure microscopique. Elle a été à la source de deux nouvelles sciences, la physique statistique et la théorie de l'information ; c'est elle qui a inspiré Shannon lorsque, comme nous l'avons vu, il a eu l'idée en 1948 d'introduire une mesure de la quantité d'information. Reliant deux concepts étrangers en apparence, elle mériterait sans doute d'être aussi célèbre que la relation d'Einstein  $E = mc^2$  qui relie énergie et masse.

### L'entropie, mesure du désordre

Introduite par Clausius à l'échelle macroscopique, l'entropie s'est donc révélée avoir une origine *microscopique et statistique*. Même si elle se manifeste à notre échelle, nous ne pouvons comprendre sa signification que dans le cadre de la physique

statistique, où elle s'identifie à une incertitude, un manque d'information : l'entropie d'un système n'est autre que la *mesure de notre méconnaissance* de ce système à l'échelle microscopique. Son expression (3) explique pourquoi elle croît (logarithmiquement) avec le volume pour un gaz : le nombre  $W$  de configurations est, pour chaque molécule, proportionnel au volume disponible. De même,  $S$  croît avec la température (ou avec l'énergie interne) parce qu'il y a de plus en plus de façons de partager l'énergie cinétique totale entre molécules lorsque cette énergie augmente.

L'entropie s'interprète aussi comme une *mesure du désordre*, notion assimilable à un manque d'information. Dire d'un paquet de cartes bien brassées qu'il est désordonné, c'est dire que l'on a perdu toute information sur leur arrangement, de sorte que l'on attribue à toutes les configurations possibles la même probabilité ; les *prévisions* que l'on peut faire sur un tirage sont aussi incertaines que possible. Au contraire, on peut prévoir avec certitude tout tirage pour un paquet de cartes classées par exemple dans l'ordre naturel du bridge ; l'entropie statistique est nulle pour cet ordre parfait. En thermodynamique, la croissance de l'ordre lorsque l'entropie diminue est parfois manifeste à notre échelle. Ainsi, lorsqu'un liquide, en se refroidissant, se transforme en cristal, il passe dans un état de la matière visiblement plus ordonné. L'augmentation d'entropie lorsque du café et du lait se mélangent traduit un accroissement du désordre dans la disposition des molécules.

Cette interprétation de l'entropie éclaire la signification microscopique du deuxième principe. Lors d'une transformation spontanée d'un système isolé, la croissance de  $S$  qui accompagne le changement des variables thermodynamiques signale une augmentation du nombre  $W$  des configurations inobservables qui leur est associé. Notre information microscopique résultante ne peut croître ; en d'autres termes, le système ne peut devenir moins désordonné qu'il n'était au départ. Ainsi, toute transformation irréversible, comme un transfert de chaleur d'un corps chaud vers un corps froid, un mélange ou une combustion, se produit dans le sens où le *désordre augmente* à l'échelle moléculaire. Une analyse de la dynamique microscopique confirme ce résultat de nature statistique. Le deuxième principe exprime donc la tendance d'un système isolé à accroître son désordre.

Lorsque plusieurs systèmes interagissent, rien m'interdit à l'un d'eux de devenir de plus en plus ordonné dès lors que le désordre total (ou l'entropie totale) croît. Dans un moteur thermique, l'énergie est fournie par la source chaude, mais en cédant de la chaleur celle-ci doit gagner du l'ordre. Une *source d'ordre* est donc indispensable : c'est le rôle que joue la source froide (tour de refroidissement dans une centrale électrique). Le désordre global augmente grâce à un flux de chaleur depuis la source chaude vers la source froide, et nous extrayons de l'énergie mécanique en détournant à notre profit une partie de ce flux grâce à un dispositif tel qu'une turbine. De même, refroidir l'intérieur d'un réfrigérateur domestique y augmente l'ordre ; mais cette augmentation doit être payée par une croissance plus forte, survenant ailleurs, du désordre ; de fait, l'électricité consommée pour faire tourner le compresseur du réfrigérateur n'est produite qu'au prix d'une augmentation de désordre dans la centrale électrique. Un autre exemple est le développement d'un être vivant, processus où l'ordre croît aux dépens du milieu environnant.

Nous avons identifié le manque d'information, grandeur subjective (ou plutôt *intersubjective* car commune à tous les observateurs disposant des mêmes informations), au désordre, grandeur en apparence objective. En réalité, l'ordre n'est pas une propriété des objets pouvant se définir de façon absolue. Un bureau considéré comme désordonné par un visiteur peut fort bien être ordonné pour son occupant s'il sait où trouver chaque document. Brassé par un prestidigitateur habile, un paquet de cartes initialement rangées peut sembler au spectateur être devenu désordonné ; cependant, le manipulateur qui a contrôlé ce brassage sait attribuer au paquet un ordre parfait, caché et connu de lui seul, de sorte qu'il est capable de restituer l'ordre initial. On a pu de même effectuer en physique des expériences de résonance

magnétique nucléaire, dites d'échos de spin, où l'entropie thermodynamique croît d'abord normalement, mais décroît ensuite en violant le deuxième principe : les moments magnétiques élémentaires, orientés parallèlement à l'origine, se désordonnent de sorte que le matériau se désaimante ; mais on peut parvenir par une manipulation macroscopique à restaurer l'ordre initial apparemment perdu, comme si on remontait le temps. L'expérimentateur est capable en pareil cas de contrôler certaines variables microscopiques habituellement inaccessibles à la thermodynamique. Dans l'état intermédiaire désaimanté, il existe un *ordre caché* caractérisé par les valeurs de ces variables microscopiques contrôlables. L'état du système doit ici être caractérisé à chaque instant non seulement par les variables thermodynamiques, mais aussi par les variables cachées contrôlables. Le nombre  $W$  correspondant est plus faible que dans une description où l'on ignore ces variables cachées, et il croît comme il se doit tout au long du processus, de même que la nouvelle entropie (3) associée à ce nombre. Bien que de telles expériences se déroulent à notre échelle, leur analyse nécessite une description plus fine que celle inhérente à la thermodynamique.

### Le démon de Maxwell

Un autre type de violation du deuxième principe se présente, du moins en apparence, dans une expérience de pensée imaginée par Maxwell en 1867. Considérons deux récipients accolés  $A$  et  $B$  séparés par une cloison percée d'un petit trou et thermiquement isolés. Ils contiennent au départ un gaz à l'équilibre, dont la densité est la même des deux côtés  $A$  et  $B$ . Un petit démon s'installe devant le trou, qu'il peut fermer à volonté à l'aide d'un opercule assez léger pour être manipulé sans apport d'énergie mécanique. En raison de l'agitation thermique, les  $N$  molécules de gaz peuvent se diriger vers le trou en venant soit d'un côté soit de l'autre. Le démon laisse passer celles qui vont de  $B$  vers  $A$ , et renvoie avec son opercule celles qui vont de  $A$  vers  $B$ . Tout le gaz finit par se retrouver dans  $A$ , de sorte que son volume s'est réduit sans compression par un piston. L'entropie du gaz a donc *décru* sans qu'il y ait eu apport de travail ni perte de chaleur, ce qui contredit le deuxième principe. Quoique cette expérience ne soit pas réalisable en pratique, l'argument de Maxwell n'en pose pas moins un sérieux problème conceptuel, et il a fallu plus d'un siècle pour élucider complètement le paradoxe.

Le progrès le plus significatif a été effectué en 1950 par Léon Brillouin. Déjà en 1929 Leo Szilard avait noté que le démon de Maxwell ne parvient à faire décroître l'entropie du gaz que grâce à des informations en sa possession : afin d'effectuer son tri, il a besoin de savoir au moins de quel côté du trou se présente chaque molécule. Peu après la naissance de la théorie de l'information, Brillouin donna une formulation mathématique à cet argument. On savait désormais chiffrer une quantité d'information disponible : par exemple, connaît de quel côté se trouve chacune des  $N$  molécules représente  $N$  bits (1 bit par molécule puisque celle-ci a une chance sur deux de se trouver en  $A$  ou en  $B$ ). Brillouin montra que la quantité d'information dont le démon doit disposer au départ pour faire décroître l'entropie d'une valeur  $\Delta S$  est au moins égale à  $\Delta S/k$ . Après le tri, le démon n'a pas plus d'information que l'observateur extérieur qui constate que tout le gaz est en  $A$ . Compte tenu de l'identification  $S = kI$  de l'entropie au désordre moléculaire (à la constante d'unités  $k$  près), l'action du démon revient donc à *transformer en ordre* dans le gaz *une information* qu'il possédait, avec des pertes éventuelles. Autrement dit, il nous a communiqué son information, au moins en partie.

Brillouin examina ensuite comment le démon avait pu acquérir l'information qu'il détenait. Pour détecter si une molécule se trouve en  $A$  ou en  $B$ , il faut par exemple l'observer optiquement. On éclaire pour cela l'intérieur du récipient à l'aide d'une source directive de lumière, et on analyse la lumière diffusée par les molécules. Cette diffusion désoriente le rayonnement et fait croître son désordre, donc son en-

tropie. Plus généralement, toute acquisition d'information, toute mesure, nécessite l'emploi d'un dispositif physique au sein duquel se produit une transformation irréversible. Le gain d'une quantité d'information  $\Delta I$  doit ainsi être payé par une *croissance*  $\Delta S$  de l'entropie de l'appareil, croissance associée à cette irréversibilité, et l'on montre que  $\Delta S$  doit être au moins égal à  $k\Delta I$ . La petitesse de  $k$  rend en général cette augmentation de l'entropie extrêmement faible. Cependant, pour le démon de Maxwell, la quantité d'information  $\Delta I$  à acquérir est énorme puisqu'elle vaut  $N$  bits. Si l'on inclut alors dans le système, à côté du gaz lui-même, les nécessaires appareillages d'observation (ainsi que le démon, éventuellement remplacé par un dispositif automatisé), le bilan global d'entropie est conforme au deuxième principe.

La petitesse de  $k$  explique aussi l'efficacité du stockage de l'information dans la matière : envoyer dans une mémoire d'ordinateur un nombre considérable de bits ne modifie guère son entropie en unités thermodynamiques.

En définitive, à la constante d'unités  $k$  près, l'identification de l'entropie à un manque d'information est complète, l'une pouvant *se transformer* en l'autre, avec des pertes éventuelles d'ordre ou d'information. Toute acquisition d'information nécessite un accroissement de l'entropie de l'appareil de mesure, d'une quantité au moins équivalente ; une information peut inversement être exploitée pour créer une quantité au plus égale de "néguentropie" (c'est-à-dire de baisse d'entropie).

## Conclusion

Au cours des dernières décennies, le concept d'entropie ou d'information a été appliqué à des domaines de plus en plus variés : théorie de la mémoire, problèmes d'optimisation, cosmologie, théorie du chaos, bio-informatique, etc. Il a aussi continué à subir diverses métamorphoses. Alors que l'entropie de la thermodynamique caractérise le désordre microscopique régnant à l'équilibre dans un système physique, l'entropie de Kolmogorov (1949) sert à préciser le caractère désordonné d'une *dynamique* et l'incertitude sur l'évolution d'un système. Par ailleurs, on a pu caractériser le désordre associé à un message ou une configuration unique, sans faire appel comme en (1) aux probabilités  $p_m$ . La quantité d'information est alors remplacée par le concept apparenté de *complexité algorithmique*, définie comme suit. On commence par numériser le message, c'est-à-dire le représenter par un nombre à l'aide d'un code. On imagine ensuite comment ce nombre peut être construit par des opérations algébriques dans un ordinateur idéal ; le logarithme de la longueur du programme informatique le plus économique pour y parvenir définit la complexité algorithmique. Pour une famille de messages, la moyenne de cette complexité s'identifie au manque d'information. Une autre direction de recherches actuelles porte sur la *théorie quantique de l'information*, en vue de l'élaboration d'hypothétiques ordinateurs quantiques ; ici, la notion de probabilité est modifiée afin de tenir compte du fait que les grandeurs physiques microscopiques sont représentées par des objets mathématiques qui ne commutent pas.

Entropie, manque d'information, incertitude, désordre, complexité, apparaissent donc comme des avatars d'un seul et même concept. Sous l'une ou l'autre de ces formes, l'entropie est associée à la notion de probabilité, et elle peut s'évaluer numériquement, ce qui en fait un outil efficace pour des applications variées. Elle caractérise non pas un objet *en soi*, mais la *connaissance* que nous en avons et nos possibilités de faire des *prévisions*. Elle a donc un caractère à la fois objectif et subjectif. Plus généralement, nous avons appris depuis un siècle que même les sciences les plus "dures", la physique et la chimie, font intervenir l'humain. Elles reposent en effet sur la mécanique quantique et sur la physique statistique, disciplines théoriques où la notion de probabilité est incontournable. Or, cette notion elle-même est à cheval sur l'objet et sur l'observateur qui cherche à le décrire et à faire des prévisions. David Ruelle, dans sa conférence donnée dans ce cycle sur



le chaos déterministe, a souligné que le hasard n'est que l'expression de nos incertitudes. Déjà Maxwell écrivait : "Confusion, like the correlative term order, is not a property of material things in themselves, but only in relation to the mind which perceives them", ce que je me permettrai de traduire par un jeu de mots : "L'entropie a un caractère anthropique".