VARIABILITÉ ET SPÉCIFICITÉ*

F. BONSACK

C'est pour moi un très grand honneur, sans doute immérité, que d'avoir été appelé à prononcer ici une conférence en séance plénière. Cet honneur, c'est avant tout à M. Couffignal que je le dois; j'aimerais l'en remercier ainsi que les autres membres du Conseil d'Administration et j'espère ne pas leur donner l'occasion de regretter leur choix.

Je vais parler de choses très élémentaires, que certains d'entre vous connaissent sans doute fort bien, et je m'en excuse auprès d'eux.

Si j'ai décidé d'en parler tout de même, c'est qu'il m'est apparu, au cours de nombreuses conversations. qu'elles ne sont pas connues de tous, même dans le monde de la cybernétique. Il ne me paraît donc pas inutile d'y revenir et d'essayer de voir clair dans certaines notions fondamentales de la théorie de l'information, notions qui ne sont pas toujours exposées avec toute la clarté désirable.

Je vais donc partir de la notion de quantité d'information selon Shannon. On peut l'introduire sommairement ainsi.

On a un *alphabet* comprenant un certain nombre de signes, à l'aide desquels on compose un *message*. On se demande comment estimer l'information contenue dans un tel message.

Il est naturel d'exiger que cette quantité d'information croisse linéairement avec le nombre de signes du message, avec sa longueur, que par exemple un message deux fois plus long contienne deux fois plus d'information.

En outre on voit bien, en essayant de traduire un message rédigé à l'aide d'un alphabet de 16 signes, dans un alphabet de 4 ou de 2 signes, que le même message requiert beaucoup plus de signes lorsqu'il est traduit dans un alphabet plus pauvre. On en déduira donc naturellement que la quantité d'information transmise par chaque signe est une fonction croissante du nombre de signes que comprend l'alphabet.

Quelle fonction ? Un alphabet de 16 lettres permet d'écrire des messages deux fois plus courts qu'un alphabet de quatre lettres, quatre fois plus courts qu'un alphabet de deux lettres. On est donc conduit au logarithme puisque $\log_2 16 = 4$, $\log_2 4 = 2$, $\log_2 = 1$, et on définira la quantité d'information par symbole

$$I_{\text{par symbole}} = \log_2 n$$

* Conférence présentée au *3e Congrès International de Cybernétique*, Namur, 11-15 septembre 1961; publiée dans *Cybernetica*, vol. IV no **3** (1961), p. 131-153.

n étant le nombre de signes que compte l'alphabet. Voilà un premier résultat, bien connu de tous.

Il faut cependant bien se rendre compte comment on l'a obtenu : on n'a pas opéré sur des messages ordinaires, rédigés dans un langage naturel, comme le français ou l'anglais. Ce serait trop compliqué. On se fait donc un modèle simplifié, où l'on considère des «messages» qui sont simplement des suites de lettres disposées plus ou moins au hasard et qui, à première vue, ne signifient rien. Mais on définit la quantité d'information comme s'il s'agissait de véritables messages, signifiant quelque chose. Cette procédure est parfaitement légitime, dans certaines limites. Mais il ne faut pas oublier que c'est sur un modèle qu'on opère, et non sur de véritables messages.

Mettons ce modèle en pleine lumière : les «messages» avec lesquels on opère sont des suites de signes et rien d'autre. Ces suites de signes sont formées par un processus aléatoire, par exemple par des jets de dés, ou par des tirages de boules dans une urne, de cartes dans un jeu brassé. À chaque signe est attribuée une certaine probabilité; si la suite est assez longue, on peut espérer, avec une certaine sécurité, que la fréquence relative des signes s'approchera de leur probabilité.

Nous allons donc considérer maintenant uniquement le modèle : une machine produisant certaines suites de signes selon certaines lois de probabilités, une machine à fabriquer des suites aléatoires de signes.

Mais ici, dans ce modèle, le terme de «quantité d'information» n'a plus aucun sens, puisqu'il n'y a plus à proprement parler d'information. Il y a une certaine grandeur du modèle qui correspond à ce que, dans des messages signifiants, on peut appeler «quantité d'information», et qui, dans le cas où tous les signes sont équiprobables, est donnée par la formule

$$H_{\text{par symbole}} = \log_2 n$$

Cette grandeur que Shannon a appelée «entropie» et qui, on ne le répétera jamais assez, n'a dans le modèle rien d'une quantité d'information, est maintenant une mesure de la *variété* des suites de signes produites par la machine-à-produire-des-suites-aléatoires-de-signes, de leur diversité, ou une mesure de la *variabilité*, de la richesse de la production de la machine, de la variabilité du «type» de «messages» produits par la machine. Plus les «messages» sont divers, variés, plus l'entropie, la variabilité est grande; plus au contraire les «messages» sont uniformes, monotones, plus leur entropie-variabilité est petite. (Pour éviter toute confusion, je précise que j'utiliserai le terme de «variabilité» lorsqu'il s'agira d'un *type de messages*, le terme de «variété» lorsque je considérerai une *collection* supposée réalisée).

Ce modèle s'avère très fécond. On peut en particulier passer à des signes de probabilités inégales; on obtient alors la fameuse formule

$$H_{\text{par symbole}} = -\sum p_i \log p_i$$

On peut aussi tenir compte des liaisons entre les signes. On se rapproche alors considérablement des langues naturelles (où les signes ne sont ni équiprobables, ni indépendants) de telle sorte qu'on pense, à première vue, pouvoir redonner à la variabilité sa signification primitive de quantité d'information.

Voilà où l'on arrive en suivant une première voie — voie que j'appellerai objective parce qu'on y examine la production réalisée, objective, d'une machine à produire des suites aléatoires de signes, production qu'on peut examiner sans se soucier de signification ou d'information proprement dite.

Mais on peut aussi suivre une voie différente, que j'appellerai subjective parce qu'elle interprète l'entropie non pas comme la variété en soi d'un certain ensemble de «messages», mais comme notre *incertitude* quant à un certain événement. On quitte le domaine de l'être pour celui de la connaissance, on passe d'une variété en soi à une variété pour nous, à la variété des messages que nous croyons pouvoir recevoir.

Cette façon d'envisager les choses est peut-être plus accessible; le modèle y est peut-être plus proche de la réalité et l'entropie peut-être plus immédiatement interprétable comme une quantité d'information.

En effet, on peut envisager des processus aléatoires — tirages de cartes ou de boules — et se poser des questions quant à l'information qu'on possède ou qu'on fournit sur le résultat de ces processus aléatoires. Par exemple, on peut effectuer un tirage de cinq boules dans une urne contenant, en quantités égales, des boules rouges et des boules blanches et se demander combien on transmet d'information sur le résultat de ce tirage en dévoilant le nombre de boules rouges tirées (sans préciser l'ordre dans lequel elles ont été tirées).

Nous reviendrons un peu plus loin sur cette conception subjective de l'information, mais je voudrais auparavant bien préciser le sens où j'entends «subjectif» et «objectif».

Il s'agit d'une subjectivité dans un sens très spécial, plus exactement d'une relativité à la situation, au point de vue. En ce sens, un aspect perspectif d'un objet est subjectif parce qu'il dépend du point de vue où l'on se place; suivant le point de vue, on verra un cercle comme un cercle ou comme une ellipse plus ou moins aplatie.

Ici, je parle de conception subjective de l'entropie parce qu'elle dépend non pas seulement des caractéristiques de l'objet, de l'événement, mais encore de ce que sait le sujet. Lorsque j'apprends quelque chose sur un objet, cet objet ne se modifie pas, en soi; mais la connaissance que j'en prends se modifie. Mon incertitude au sujet d'un événement peut se modifier; elle peut diminuer si l'on me fournit des informations; l'événement se modifie alors *pour moi*, mais non pas *en soi*; c'est une modification que j'appelle donc subjective.

Elle est encore subjective en ce sens qu'elle varie suivant ce que sait tel ou tel sujet. Supposons, par exemple, qu'on examine l'information apportée par une certaine nouvelle. D'après la définition subjective, la quantité d'information n'est pas la même pour un récepteur qui ne l'attend pas, qui est dans une ignorance, dans une incertitude totale quant à cette nouvelle, que pour un récepteur qui soupçonne déjà quelque chose de précis, qui a donc avant la réception de l'information une incertitude moins grande. Pour un récepteur qui connaît déjà la nouvelle, celle-ci n'apporte plus d'information, car il n'y a déjà plus aucune incertitude avant la réception de l'information.

Il y a donc une certaine subjectivité, en ce sens que la quantité d'information dépend de la situation de connaissance dans laquelle se trouvait le récepteur avant de recevoir la nouvelle; mais cette subjectivité n'a rien de la subjectivité du poète ou du passionné; étant donnée telle situation de connaissance, l'information est univoquement déterminée; elle sera la même pour tous les individus dans la même situation, quels que soient leurs goûts ou leur tempérament. Elle est moins subjective au sens courant que *relative à la situation*.

Je ne vais pas choisir entre la conception objective et la conception subjective de l'information. Personnellement, je trouve la conception objective plus satisfaisante et plus générale; il est en particulier très facile de retrouver la conception subjective en partant de la conception objective. D'autre part, la conception objective ouvre des perspectives du côté de ce qu'on a quelquefois appelé la «néguentropie structurale», l'entropie d'un certain type de structures qui n'ont rien d'informationnel. Mais je dois reconnaître que la conception subjective présente certains avantages d'exposition et peut-être moins de dangers de confusions.

L'erreur fréquente que je voudrais ici dénoncer n'est d'ailleurs pas particulière à la conception subjective ou objective : on la retrouve dans les deux conceptions. Elle consiste à identifier simplement la quantité d'information à l'entropie, à définir la quantité d'information par exemple à l'aide de la formule $-\sum p_i \log p_i$ et à s'en tenir là. Ce n'est pas seulement faux parce que l'entropie est une grandeur du modèle alors que la quantité d'information est une grandeur du domaine représenté — car il pourrait y avoir correspondance assez rigoureuse entre le modèle et ce qu'il représente. C'est plus grave : l'entropie n'est pas la grandeur du modèle qui correspond à la quantité d'information, c'est une autre grandeur, très proche, mais non identique, qui lui correspond. Shannon ne commet pas d'erreur dans son exposé, mais il ne souligne nulle part qu'après le premier chapitre, l'entropie cesse d'être une mesure de la quantité d'information, bien qu'il utilise lui-même quelque chose d'autre.

Si l'on n'est pas suffisamment conscient de ce fait, on risque de se heurter à des paradoxes que l'on n'arrive pas à débrouiller. J'en parle en connaissance de cause, puisque j'ai buté pendant plusieurs années sur de tels paradoxes, avant de découvrir où était l'erreur. C'est dans l'espoir d'éviter de pareils écueils à quelques-uns que j'ai choisi de traiter ce sujet devant vous.

Ces paradoxes se manifestent surtout lorsqu'on aborde le problème de la transmission avec bruit. En voici un exemple.

Prenons des messages formés de deux signes, 0 et 1. Ces deux signes ne sont pas équiprobables : 0 a la probabilité 3/4 et 1 la probabilité 1/4.

Calculons l'entropie d'un tel message à l'aide de la formule

$$H_{\text{par symbole}} = -\sum p_i \log p_i$$

On a

$$H_{\text{par symbole}} = -(3/4 \log_2 3/4 + 1/4 \log_2 1/4) = 0.811$$

Introduisons un bruit, par exemple 1/4 d'erreurs (une erreur étant un 0 pour un 1 ou un 1 pour un 0) à la transmission. On a le schéma suivant

$$\begin{array}{ll} 3/4\ 0 & \left\langle \begin{array}{ll} 3/4 \text{ corr.} & 9/16\ 0 \\ 1/4 \text{ err.} & 3/16\ 1 \end{array}\right. \\ 1/4\ 1 & \left\langle \begin{array}{ll} 3/16 \text{ corr.} & 3/16\ 1 \\ 1/16 \text{ err.} & 1/16\ 0 \end{array}\right. \end{array}$$

On a donc, après perturbation, 9/16 + 1/16 = 5/8 de 0 et 3/16 + 3/16 = 3/8 de 1. L'entropie du message perturbé est

$$H_{\text{par symbole}} = -(3/8 \log_2 3/8 + 5/8 \log_2 5/8) = 0.955$$

L'entropie est donc passée de 0.811 à 0.955; elle a augmenté. C'est tout à fait normal si l'on interprète l'entropie comme une variabilité : le bruit a introduit une nouvelle source de variabilité, on s'attend donc que celle-ci ait augmenté.

Mais la quantité d'information, elle, a bien évidemment diminué : le bruit, les perturbations n'augmentent pas la quantité d'information, bien au contraire.

Que conclure de cette contradiction ?

On ne peut en conclure qu'une chose, c'est que l'entropie n'est pas la grandeur qui doit représenter la quantité d'information. Il faut donc se mettre à la recherche d'une autre grandeur, qui ne nous mène pas à de semblables paradoxes mais qui explique cependant qu'on ait pu prendre tout d'abord l'entropie-variabilité pour une bonne mesure de la quantité d'information.

Que fait l'expéditeur d'un message ?

Il a à sa disposition une collection de messages possibles d'une certaine étendue ou richesse mesurée par une certaine entropie-variabilité. Par exemple, cette collection sera l'ensemble des suites de signes d'une certaine longueur qu'on peut composer à l'aide des touches d'un téléscripteur.

Mais il ne veut pas envoyer n'importe quel message, n'importe quelle suite de lettres. Il choisit *un message particulier* dans cette collection; c'est ce message particulier qu'il veut envoyer, et aucun autre. C'est d'autre part ce message particulier et aucun autre qui transmettra telles informations au destinataire, informations qu'il désire avoir ou que l'expéditeur désire lui communiquer.

Il semble donc que l'information liée à un message soit liée à sa spécificité, au fait qu'il est ce message et aucun autre. Et si un brouillage en fait un message quelconque, le rend méconnaissable, il n'apportera plus aucune information, même si son entropie-variabilité est toujours très élevée.

Essayons de voir cela un peu plus clairement sur un exemple particulier.

Prenons des messages de 2 signes, 0 et 1, équiprobables. Transmettons-les avec 1/4 d'erreurs. Leur entropie-variabilité n'aura pas changé puisqu'il y aura autant de 0 qui deviendront des 1 que de 1 qui deviendront des 0. On aura donc à la réception de nouveau des messages composés de 0 et de 1, équiprobables, et qui ne se distingueront en rien des messages non perturbés. Mais quelque chose a changé, quelque chose à quoi doit être liée la quantité d'information : ce ne peut être que le fait que la spécificité des messages choisis par l'expéditeur a diminué.

Supposons que l'expéditeur transmette un grand nombre de fois le même message et examinons ce que ce message est devenu après transmission. Si la transmission se fait sans bruit, ce message donnera toujours le même message : la spécificité du message et l'information qu'il transmet éventuellement sont donc conservés. Mais si la transmission est perturbée, ce même message donnera naissance, à la réception, à une collection de messages d'une certaine variété : dans l'un, il y aura une erreur à la première lettre et à la cinquième, dans l'autre, à la troisième seulement, etc. Plus les perturbations seront nombreuses, plus la variété de cette collection augmentera et moins grande sera la quantité d'information réellement transmise.

Examinons encore un autre cas. Supposons que j'aie convenu avec mon correspondant que seule une lettre sur deux soit significative, que le reste soit un remplissage quelconque. Dans ce cas, à mon message significatif correspondra un grand nombre de messages différant par le remplissage, mais identiques pour les lettres significatives. Ici encore, la quantité d'information par symbole est nettement moins grande que si tous les symboles avaient été significatifs, et en même temps la spécificité du message a diminué, puisque le message unique a été remplacé par toute une collection de messages équivalents.

Ces considérations suggèrent une solution à notre paradoxe.

On a deux collections : la collection de messages dont on dispose avant le choix et la collection des messages qu'on a sélectionnés parmi ceux de la première collection. Plus la première collection est étendue, diverse, plus un message choisi pourra contenir d'information; cela, nous l'avons déjà établi dans la première partie. La quantité d'information est donc une fonction croissante de l'étendue de la collection avant le choix. Mais c'est une fonction décroissante de l'étendue de la collection choisie : plus cette collection est large, moins le choix est spécifique, moins grande est la quantité d'information. Nous allons mesurer l'étendue de ces deux collections par leur entropie-variété et définir la *spécificité* de la collection choisie — cette spécificité qui est la mesure de la quantité d'information — par la différence entre ces deux entropies

$$\text{Spécificité} = H_0 - H_1$$

H_0 étant l'entropie de la collection offerte à l'expéditeur et H_1 étant l'entropie-variété de la collection choisie.

La spécificité est donc une mesure de la particularité, de l'originalité d'une certaine sous-classe restreinte par rapport à une classe plus large, à un référentiel. Cette spécificité est d'autant plus grande que la sous-classe est plus restreinte ou que le référentiel est plus étendu. Par exemple, la sous-classe des individus à yeux bleus a une certaine spécificité par rapport au référentiel des hommes en général; elle a une spécificité moins grande par rapport aux habitants des pays nordiques; par contre, la sous-classe des individus à yeux bleus et à cheveux bruns a une spécificité plus grande que celle groupant tous les individus ayant les yeux bleus, quelle que soit la couleur des cheveux.

Cette notion de spécificité permettra-t-elle de lever notre paradoxe ? Mènera-t-elle à des résultats raisonnables ?

Essayons de l'appliquer à la transmission avec bruit. L'expéditeur avait à sa disposition une certaine collection de messages d'entropie-variété H_0. Parmi ces messages, il en choisit un seul. La collection n'a plus aucune variété; le nombre de possibilités équiprobables étant égal à 1, l'entropie, qui s'obtient en prenant le logarithme de ce nombre de possibilités, est égal à 0; H_1 est donc ici nulle. Dans ce cas, l'entropie-variété de la collection offerte à l'expéditeur est donc bien une mesure de la quantité d'information, puisqu'on n'a rien à lui soustraire.

Mais, après perturbation, l'entropie-variabilité du message ne sera plus nulle; à un certain message émis correspondra une certaine collection de messages possibles à la réception. On doit donc estimer la spécificité de cette collection correspondant à un message particulier par rapport à l'ensemble des messages possibles à la réception (correspondant à des messages quelconques à l'émission). En d'autres termes, on cherche à calculer la spécificité de ce qu'est devenu le message par rapport à l'ensemble des messages possibles à la réception.

Dans l'exemple de tout à l'heure (3/4 de 0 pour 1/4 de 1, 1/4 d'erreurs à la transmission), nous avions calculé l'entropie des messages à la réception. Elle était égale à 0.955. Nous devons soustraire de cette entropie, l'entropie de la collection issue d'un message particulier; cette entropie est égale à

$$-\,(3/4 \log_2 3/4 + 1/4 \log_2 1/4) = 0.811$$

La spécificité de cette collection par rapport à l'ensemble des messages possibles à la réception est donc égale à

$$0.955 - 0.811 = 0.144 \text{ bit par symbole}$$

Un quart d'erreurs fait donc perdre plus des 4/5 de l'information, ce qui peut paraître surprenant à première vue, mais s'explique par le fait qu'on perd non seulement de l'information parce que telle lettre a été substituée à telle autre, mais encore parce qu'on ignore où ces substitutions ont été faites, parce que rien ne distingue un *a* issu d'un *a* d'un *a* issu d'un *e* ou de n'importe quelle autre lettre. En supprimant simplement des lettres et en signalant l'endroit où une lettre a été supprimée, 1/4 de suppressions ne diminuerait l'information que d'un quart.

Passons à l'autre exemple, celui où les 0 et les 1 sont équiprobables; l'entropie-variabilité par signe est donc égale à 1.

Introduisons un nombre considérable d'erreurs : supposons que la moitié des signes soient perturbés.

Nous avons déjà vu que l'entropie-variabilité était la même à l'émission et à la réception. Quelle est la spécificité des messages reçus ? La variété de la collection issue d'un message donné vaut

$$-(1/2 \log_2 1/2 + 1/2 \log_2 1/2) = 1$$

L'entropie-variété de cette collection est égale à celle de tous les messages à la réception. La spécificité est donc nulle, et la quantité d'information aussi : en toute rigueur, plus aucune information n'est transmise. Il y a bien encore la moitié des signes qui sont corrects, mais comme rien ne permet plus de les reconnaître, ils ne peuvent être utilisés. On obtiendrait une aussi bonne «transmission» en coupant la ligne et en composant le message en jouant à pile ou face; dans ce cas aussi, il y a en moyenne un signe sur deux qui coïncide avec le signe correspondant de n'importe quel original.

Cet exemple nous permet de préciser encore un peu le sens du mot spécificité. Tant que la collection des messages issus d'un message particulier est plus restreinte que la collection de tous les messages possibles à la réception, cette collection garde une certaine particularité, une certaine spécificité; les collections issues de messages différents se distinguent encore les unes des autres, il y a encore des traces, des restes de la structure du message émis. Mais lorsque le message a été complètement désorganisé, la collection issue d'un message particulier n'a plus rien de particulier, plus rien de spécifique, plus rien de typique; elle ne se distingue plus de la collection issue de n'importe quel autre message, toutes les traces de la structure primitive ont disparu, la spécificité tombe à zéro. On peut caractériser brièvement ce fait en disant que le désordre complet n'a pas d'odeur, qu'il ne trahit plus rien de son origine, parce que tous les désordres absolus se ressemblent.

On arrive d'ailleurs exactement à la même formule, pour la quantité d'information, dans l'interprétation subjective. L'entropie y est interprétée comme une incertitude quant au message. Si la transmission est absolument fidèle, on passe de l'incertitude maximale à une incertitude nulle; l'information apportée par le message est donc égale à l'incertitude au départ.

Mais si le message est brouillé, sa réception diminue bien un peu l'incertitude, mais ne la supprime pas totalement. La quantité d'information apportée par le message est alors égale à la diminution d'incertitude qu'il a permise, à la différence entre l'entropie avant et l'entropie après la réception du message

$$I = H_0 - H_1$$

Si l'incertitude est restée ce qu'elle était, l'information reçue est nulle. Tout ceci s'accorde fort bien avec la notion intuitive d'information.

Arrêtons-nous un instant pour faire le point.

Nous avons défini :

- d'une part, une *entropie*, qui est une mesure de la *variabilité* d'un certain type de messages, ou de la *variété* de la collection de messages de même type à laquelle il appartient;
- d'autre part, une *spécificité*, qui est une mesure de la particularité, de l'originalité d'un message ou d'un certain groupe de messages par rapport à l'ensemble des messages de même type. Cette spécificité s'exprime par une différence entre deux entropies-variabilités. C'est cette spécificité, et non l'entropie-variabilité, qui représente dans le modèle la quantité d'information.

Quels sont les rapports, les ressemblances et les différences entre ces deux notions ?

L'un des rapports est évident : la spécificité est une différence entre deux variabilités.

Nous avons déjà fait allusion à un autre rapport : lorsque la sous-classe dont on estime la spécificité ne contient plus qu'un seul élément, la spécificité de cet élément (exactement de ce type d'élément singulier) est égale à l'entropie-variabilité du référentiel (puisqu'on lui soustrait une entropie nulle).

Mais il y a aussi des différences.

La différence essentielle, c'est que la spécificité est relative, alors que l'entropie-variabilité est absolue. La variabilité comporte une échelle absolue, du fait qu'on n'a besoin d'aucune convention, d'aucun choix de référence pour situer le zéro de cette échelle : il y a entropie ou variabilité nulle lorsqu'il n'y a qu'une possibilité, lorsque le type ne comprend qu'une seule forme ou, dans l'interprétation subjective, s'il y a certitude. La spécificité est au contraire relative; il faut toujours préciser par rapport à quel référentiel on la mesure. Un message est moins spécifique par rapport à un référentiel lui-même hautement spécifique que par rapport à un référentiel peu spécifique. Un texte français par exemple aura une spécificité moins grande par rapport à l'ensemble des textes français que par rapport à l'ensemble de tous les textes possibles dans toutes les langues, ou par rapport à des suites aléatoires de signes sans aucune liaison.

Cette relativité de la spécificité, le fait qu'elle soit définie comme une différence de deux termes, la rapproche d'une part d'une différence de potentiel, comme le font fort justement remarquer les Yaglom dans leur petit livre sur *la probabilité et l'information*, d'autre part, de la *probabilité*, qui est elle aussi relative; elle est constituée par un rapport entre deux ensembles d'événements, l'ensemble des cas favorables et celui de tous les cas possibles. Ce rapport, ce quotient est devenu ici une différence, parce qu'on est passé des nombres aux logarithmes.

Il faut également s'arrêter un instant à la question du signe; certains auteurs, M. Brillouin en particulier, ont identifié la quantité d'information à une entropie négative, à une néguentropie.

Cette solution me paraît un peu trop sommaire. Le problème se présente en réalité de la manière suivante.

Tout d'abord, nous le verrons tout à l'heure, la définition — dans la théorie moléculaire — de l'entropie a subi au cours de son histoire une profonde modification : on est passé d'une définition probabiliste selon Boltzmann à une définition statistique selon Planck; au lieu de définir, comme Boltzmann, l'entropie à partir du logarithme d'une probabilité, on l'a définie à partir du logarithme d'un nombre de complexions. Ce faisant, on a, à mon avis, inversé le signe de la relation de l'entropie avec la probabilité, car le nombre de complexions, c'est l'inverse de la probabilité d'une complexion, si toutes les complexions sont équiprobables. L'entropie de Shannon est définie, selon Planck, à partir du nombre de complexions (ou, si l'on veut, à partir du logarithme négatif d'une probabilité). Il y a déjà là une question de signe à laquelle il faut prendre garde.

Mais il y en a une autre.

La question des signes se pose surtout à propos de la variation d'entropie; ce qui importe, c'est de savoir si l'entropie augmente ou diminue.

Or il semble bien que l'entropie augmente lorsque la quantité d'information diminue — c'est ce qui a amené M. Brillouin à son identification de la quantité d'information à une néguentropie. Comment la chose se présente-t-elle si l'on introduit la notion de spécificité ?

La spécificité comprend deux termes : l'un positif, l'entropie du référentiel; l'autre négatif, l'entropie du sous-ensemble dont on estime la spécificité.

La réponse est donc très claire : si la variation d'entropie touche le premier terme, il n'y a aucune inversion de signe; une augmentation de l'entropie du référentiel augmente la spécificité. Mais si la variation touche le second terme, il y a alors inversion du signe de cette variation; si l'entropie du sous-ensemble sélectionné augmente, la spécificité diminue. Je crois que cette façon de présenter les choses est plus claire et plus précise que celle qui consiste à identifier simplement information et néguentropie.

Nous en avons ainsi provisoirement terminé avec la théorie de l'information et nous allons passer à la thermodynamique pour voir si, là aussi, les notions de variabilité et de spécificité sont utilisables et si elles apportent quelque clarté.

La notion d'entropie en thermodynamique n'a jamais été très facile d'accès. Poincaré disait d'elle que c'est une notion prodigieusement abstraite.

On a essayé, en vain, d'en donner des illustrations intuitives, des modèles mécaniques. Ostwald a imaginé une ingénieuse construction où toutes les formes d'énergie sont le produit de deux grandeurs : une extensité — par exemple la

masse, ou le volume, ou la charge — et une intensité — par exemple le carré de la vitesse, la pression ou la différence de potentiel. L'énergie calorifique se ramènerait de même à un produit de l'extensité entropie par l'intensité température. Mais tout ceci n'est pas très éclairant, et c'est peu satisfaisant, car l'entropie n'est pas une grandeur conservative, contrairement à d'autres extensités comme la masse, ou la charge. (On m'objectera peut-être que le volume n'est pas non plus une grandeur conservative au sens strict, puisqu'on peut faire varier le volume d'un gaz alors que sa pression diminue et que son énergie reste constante. Il y a cependant une différence fondamentale entre le volume et l'entropie. Si l'on maintient la pression constante, on ne peut faire augmenter le volume et l'énergie qu'en introduisant quelque chose dans le système, ici, de la matière, du gaz par exemple. Par contre, l'entropie d'un système peut augmenter alors que ce système est rigoureusement fermé, que son énergie et sa température restent constantes. L'analogie avec les autres extensités n'est légitime que si l'on se limite aux processus réversibles, où l'entropie se conserve).

Boltzmann a alors proposé une interprétation de l'entropie dans le cadre de la théorie cinétique.

Examinons l'exemple suivant.

On considère deux récipients, l'un plein d'un gaz parfait, l'autre vide. Et on les met en communication. Selon la thermodynamique classique, l'entropie augmente.

Boltzmann divise le volume des deux récipients en petites cellules de même grandeur. Et il étudie la répartition des molécules dans ces cellules. Il montre alors que la répartition la plus probable est celle où il y a le même nombre de molécules dans toutes les cellules, et que l'énorme majorité des complexions comporte une répartition *à peu près* égale.

Boltzmann met donc l'entropie en rapport avec la probabilité : dans le processus de la diffusion d'un gaz parfait, le système passe d'un état de répartition inégale, où les cellules de l'un des récipients sont toutes pleines et celles de l'autre toutes vides, à un état de répartition à peu près égale dans toutes les cellules des deux récipients. La première répartition est très improbable, la seconde est très probable. Or la première a une entropie moins élevée que la seconde. Boltzmann a donc mis en relation l'entropie et la probabilité selon la formule

$$S = k \ln W$$

(plus exactement, c'est Planck qui a donné cette forme à la relation de Boltzmann).

Il en est de même pour la température : on divise l'espace des moments en cellules et on examine la répartition des molécules dans ces cellules (les molécules groupées dans la même cellule ayant à peu près la même direction et la même vitesse, si elles ont toutes la même masse). On montre de la même manière que, pour le volume, les complexions les plus probables correspondent à la répartition égale des molécules dans les cellules et qu'une égalisation de température entre un corps chaud et un corps froid fait passer le système des deux corps d'un état très improbable vers un état très probable en même temps que l'entropie s'accroît.

Il est évident que cette définition probabiliste de l'entropie contient une part d'arbitraire. D'une part, le volume des cellules de volume ou des cellules des moments n'est pas fixé, et la probabilité en dépend. D'autre part, on peut se demander ce que veulent dire «probable» et «improbable» dans l'optique de Boltzmann. L'état de la situation initiale (où la répartition est inégale) n'est pas probable ou improbable en soi; si les récipients sont séparés, le système se trouve dans un état de probabilité maximum compatible avec cette situation. La situation initiale n'est improbable que dans la situation finale : il est extrêmement improbable que, *les deux récipients étant en communication* et les pressions ayant eu le temps de s'égaliser, toutes les molécules se trouvent par hasard dans l'un des deux récipients.

On voit immédiatement que l'entropie de Boltzmann est une entropie *relative*, qu'il s'agit de l'entropie de l'état initial par rapport à l'état final ou de l'état final par rapport à l'état initial, en un mot de l'augmentation de l'entropie entre l'état initial et l'état final.

Ceci nous conduit à interpréter l'entropie de Boltzmann comme une *spécificité* — puisqu'elle est relative et égale à une différence entre deux entropies; il s'agit donc de la spécificité de l'état initial par rapport à l'état final. Mais il y a encore une difficulté, c'est que — si l'on adopte les conventions de signes de Clausius — l'entropie *augmente* de l'état initial à l'état final. Or la spécificité de l'état initial a une certaine valeur positive, alors que la spécificité de l'état final par rapport à ce même état final est évidemment nulle; la spécificité *diminue* donc. Il faut par conséquent inverser le signe de la spécificité : *l'entropie de Boltzmann est une spécificité négative*.

Dans cette optique boltzmannienne, l'interprétation de la quantité d'information — mesurée par la spécificité — comme une entropie négative paraît donc tout à fait justifiée.

Mais l'histoire ne s'est pas arrêtée là. Clausius avait certes commencé par définir la variation d'entropie relative à l'aide de la formule

$$dS = \frac{dQ}{T}$$

Mais il a rapidement été amené à postuler des entropies absolues, uniquement fonction de l'état du système. Il y a été amené avant tout par le fait que l'augmentation de l'entropie se présente comme une différentielle totale, ce qui signifie qu'elle ne dépend que de l'état initial et de l'état final et non du chemin parcouru (à condition que ce chemin utilise des processus réversibles). De ce fait, il suffit de fixer arbitrairement l'entropie d'un seul état pour que l'entropie devienne une fonction univoque d'un état quelconque, un potentiel, de même qu'il suffit de fixer

arbitrairement une altitude — par exemple celle du niveau de la mer, qu'on définira comme étant l'altitude zéro — pour que l'altitude devienne une fonction univoque de chaque point du terrain. Il intervenait bien, dans cette définition de l'entropie absolue, une constante additive arbitraire, non déterminable dans le cadre de la thermodynamique classique, mais cette constante arbitraire n'était pas gênante, puisqu'il suffisait de la fixer une fois pour toutes.

Puis est venu Nernst, avec son «troisième principe de la thermodynamique»; Nernst a établi que, lorsqu'on s'approchait du zéro absolu, l'entropie tendait vers une limite inférieure qu'elle ne pouvait dépasser.

Planck a alors fait une proposition fort raisonnable : puisqu'il suffit, pour définir l'entropie absolue d'un état quelconque E, de fixer arbitrairement l'entropie E_0 d'un état particulier, puisque d'autre part on sait que l'entropie s'approche d'une limite inférieure au zéro absolu, nous allons tout arranger de façon très commode en fixant à zéro l'entropie d'un système se trouvant au zéro absolu. Cela fera en outre une belle symétrie entre l'entropie et la température.

On s'est donc résolument engagé dans la définition d'entropies absolues, tout d'abord pour des raisons de commodité, semble-t-il.

Mais ainsi, on s'éloignait de l'interprétation de Boltzmann, qui faisait de l'entropie une spécificité négative. On se rapprochait au contraire d'une interprétation de l'entropie comme une variabilité.

C'est encore Planck qui a franchi le pas décisif. On éprouvait, nous l'avons vu, certaines difficultés à fixer la probabilité W de Boltzmann, qui restait relative d'une part au volume des cellules choisi, d'autre part à l'état final pris comme référentiel. La mécanique ondulatoire a permis de fixer une valeur définie pour le produit du volume des cellules de volume par celui des cellules de l'espace des moments. (C'est l'une des formes qu'on peut donner au principe d'incertitude de Heisenberg : si l'on réduit les dimensions des cellules de volume, on accroît la précision de la localisation des molécules, mais on augmente en même temps l'incertitude sur leur moment, c'est-à-dire qu'on ne peut plus répartir les molécules que dans de grandes cellules de l'espace des moments). Ceci fixait univoquement le nombre de répartitions possibles dans telle situation expérimentale macroscopique, dans tel état macroscopique. Autrement dit, la probabilité de Boltzmann devenait un rapport entre deux nombres ne dépendant chacun que de l'état correspondant, de même que l'entropie relative, la variation d'entropie, était donnée par la différence des entropies absolues (le passage du rapport à la différence s'explique ici encore par l'intervention du logarithme). Tout invitait donc à faire correspondre à l'entropie absolue le logarithme du nombre de complexions P

$$S = k \ln P$$

C'est précisément ce qu'a fait Planck, substituant à la définition probabiliste de Boltzmann une définition *statistique*, ne faisant intervenir que le nombre de complexions.

Mais cette définition correspond parfaitement, à une constante près, à la définition que nous avions donnée de l'entropie-variabilité dans le cas où les signes sont équiprobables. L'interprétation de Planck est donc une interprétation en variabilité, ce qui justifie l'emploi, par Shannon, du terme d'entropie pour la variabilité en théorie de l'information.

Que signifie cette nouvelle interprétation ? À quoi correspond, que représente maintenant l'entropie thermodynamique ?

L'entropie d'un état macroscopique mesure la diversité, la variété des états microscopiques compatibles avec cet état macroscopique. Par «état microscopique», nous entendons un état défini par la répartition des molécules dans les cellules du volume et de l'espace des moments. Par «état macroscopique», nous entendons un état défini par des grandeurs macroscopiques telles que masse, volume, pression, température.

Avec cette interprétation, l'entropie cesse d'être la notion prodigieusement abstraite qu'elle était en thermodynamique classique et la notion assez nébuleuse qu'elle était dans la théorie de Boltzmann. Ses caractères, ses particularités deviennent tout à fait compréhensibles.

Par exemple : lorsqu'un gaz parfait se détend dans le vide, ou lorsqu'il augmente de volume sans fournir de travail, son entropie augmente. Comment cela s'interprète-t-il ?

C'est tout simple : on lui offre de nouvelles cellules pour ses molécules. Le nombre de répartitions possibles va donc augmenter, et avec lui l'entropie-variabilité microscopique du nouvel état. Un exemple élémentaire le montrera bien. Supposons que nous ayons deux molécules et deux cellules; le nombre de complexions possibles est trois :

	o o		o	o		o o
Probabilité :	1/4		1/2		1/4	

L'entropie, calculée en bits, est égale à 1.5.

Mais augmentons le volume en ajoutant une cellule. Le nombre de complexions passe alors à 6 :

	o o			o	o		o		o
Prob.:		1/9			2/9			2/9	
		o o			o	o			o o
		1/9			2/9			1/9	

L'entropie de ce nouvel état est de 2.507 bits. Elle a donc bien augmenté.

Il en est de même pour la température : une augmentation de la température ne correspond pas seulement à une augmentation de l'énergie cinétique moyenne des molécules, mais à une augmentation de la diversité de leurs vitesses, en valeur absolue et en direction. C'est pourquoi l'entropie du système augmente.

Enfin, le principe de Nernst s'interprète très naturellement : si l'entropie absolue est nulle au zéro absolu, cela signifie que la variété des états microscopiques est nulle, donc que le système présente toujours la même répartition de ses molécules. Et c'est évident, du moins pour la température : si toutes les molécules ont la vitesse zéro, leurs vitesses n'auront aucune variété, toutes les molécules se trouvant toujours dans la même cellule autour de l'origine, aussi petite qu'on choisisse cette cellule. Pour le volume, c'est un peu moins évident; mais du fait que le produit des cellules de l'espace des moments par les cellules de volume est fixé, si l'on peut réduire la cellule des moments autant qu'on le désire, on peut augmenter la cellule de volume autant qu'on le désire et la faire, par exemple, coïncider avec le volume occupé par le gaz. Il est clair que là encore il n'y a qu'une répartition possible, puisqu'il n'y a qu'une seule cellule.

Bref, tout ceci me paraît assez satisfaisant : les notions de variabilité et de spécificité ont permis d'éclairer l'histoire de la notion d'entropie et son interprétation actuelle.

Nous pouvons maintenant aborder un dernier chapitre qui pourrait s'intituler : *Les principes de Carnot*.

Nous n'examinerons pas le principe de Carnot et son rôle en thermodynamique classique. Ce qui nous intéresse, c'est sa traduction en théorie cinétique, avec les interprétations de Boltzmann et de Planck.

Le principe de Carnot peut se formuler ainsi : *L'entropie d'un système fermé ne peut qu'augmenter ou, à la limite, rester constante dans des processus réversibles idéaux; jamais elle ne peut diminuer*.

Comment interpréter ceci en théorie cinétique ?

Si l'on accepte l'interprétation de Boltzmann, en spécificité négative, on obtiendra un principe qui dira à peu près ceci : *Un système fermé évolue vers des états de plus en plus probables, ou de moins en moins spécifiques*.

Si, au contraire, on adopte la solution de Planck, on aura : *Un système fermé évolue vers des états de plus en plus variables*, «plus variables», prenant ici le sens particulier de «ayant une entropie-variabilité plus grande».

À première vue, la première formule, correspondant à l'interprétation de Boltzmann, paraît plus évidente et, par conséquent, préférable : il paraît évident qu'un système évolue vers des états probables; il l'est beaucoup moins qu'il évolue vers des états variables.

Mais il faut se méfier de certaines évidences.

Que signifie «probable» dans le sens de Boltzmann ? Nous l'avons vu : à chaque état macroscopique, à l'état initial aussi bien qu'à l'état final, correspond un certain nombre d'états microscopiques possibles. On se place dans la perspective de l'état final : les complexions correspondant à l'état initial sont improbables parmi toutes les complexions possibles dans l'état final, c'est-à-dire que le nombre de complexions a augmenté de l'état initial à l'état final. C'est précisément ce que dit la seconde formule, dans l'interprétation selon Planck. Les deux formules sont donc équivalentes, si l'on examine des choses avec soin.

Mais pourquoi la première paraît-elle plus évidente ? Elle est évidente si on lui donne le sens suivant : *Dans n'importe quelle situation expérimentale, le système se trouvera probablement dans un état probable*. Ce principe est évident, mais il ne dit pas assez. Il assure bien une évolution vers un état probable — encore faut-il pour cela négliger les fluctuations qui feraient revenir le système à un état improbable — mais il ne précise pas que l'état initial est toujours improbable par rapport à l'état final. Et c'est finalement ce qui est important.

La partie du principe qui est évidente est donc insuffisante et l'autre, qui est indispensable, se ramène à la seconde formule, de telle sorte que c'est cette seconde formule qui est finalement la plus simple et la plus directe.

Nous devons maintenant poser le problème de l'application du principe de Carnot à la théorie de l'information et éventuellement celui de l'établissement d'un «principe de Carnot généralisé», valable à la fois pour la thermodynamique et la théorie de l'information.

Je n'insisterai pas sur les difficultés qui apparaissent lorsqu'on veut faire de la quantité d'information une variabilité. En particulier, Shannon a démontré un théorème n° 7 dans lequel il établit que la variabilité d'un message ne peut que rester constante ou diminuer lors de son passage à travers un traducteur «déterministe», qu'en aucun cas elle ne peut augmenter. Malgré son analogie avec le principe de Carnot — analogie qui a séduit M. Brillouin — ce théorème dit exactement le contraire : l'évolution autorisée par le principe de Carnot correspond à une *augmentation* de la variabilité et non à une diminution, comme c'est le cas ici. Au contraire, la transmission avec bruit — touchée par le théorème n° 7 de Shannon — provoque en général, si la perturbation est vraiment aléatoire, une augmentation de la variabilité. On voit donc qu'il règne une extrême confusion dans les principes de Carnot de la théorie de l'information tant qu'on veut faire de la quantité d'information une variabilité.

Tout se simplifie au contraire lorsqu'on fait de celle-ci une spécificité — comme on doit le faire. Aussi bien la traduction déterministe que la perturbation aléatoire provoquent alors une diminution de la spécificité; le principe devient alors : *Un message évolue vers des états de moins en moins spécifiques* et il est identique au principe de Carnot de la thermodynamique selon Boltzmann. Le problème d'un «principe de Carnot généralisé» semble donc avoir reçu une solution dans cette perspective.

Mais ce principe pose des problèmes. En particulier, il semble être contredit chaque fois qu'il y a sélection, choix, tri. Car alors, la spécificité augmente. Il faut expliquer comment cela est possible.

M. Brillouin l'explique ainsi : pour qu'un tri soit possible, il faut que celui qui trie acquière de l'information sur ce qu'il trie, il faut qu'il sache si la boule qu'il tient en ce moment est rouge ou blanche, pour savoir où il doit la mettre. Cette information correspond, nous l'avons vu, à une certaine spécificité. Pour acquérir cette information, il doit *voir* ce qu'il trie, et pour le voir, il doit s'éclairer. Or il ne peut s'éclairer efficacement sans utiliser des processus augmentant l'entropie, donc diminuant la spécificité. Ce qu'il gagne en spécificité macroscopique, par le tri, il ne peut le gagner qu'en acquérant de la spécificité informationnelle, cette spécificité informationnelle ne pouvant à son tour être acquise qu'au prix d'une dépense de spécificité physique. Il y aurait donc «compensation», et c'est là un phénomène bien connu en thermodynamique : l'entropie peut bien décroître dans un sous-système ouvert, mais elle croît alors dans une autre partie du système fermé total. Par exemple une machine frigorifique est capable d'abaisser la température d'un corps aux dépens de l'élévation de température d'un autre corps, ce qui représente une diminution d'entropie. Mais cette diminution se paye par une dépense de travail, dépense représentant une augmentation d'entropie qui compense et même surcompense en général la diminution obtenue. M. Brillouin postule qu'il en est de même pour la diminution d'entropie provoquée par le tri.

Je suis parfaitement d'accord sur la compensation. Mais je le suis moins sur la manière dont elle est justifiée. En effet, il peut y avoir tri, sélection, sans qu'il y ait information : si l'on passe des cailloux à travers un crible, on sélectionne les plus fins; il n'y a là aucune information, aucune «torche éclairante». Et pourtant, la spécificité augmente. (On m'a objecté qu'il y avait dans le crible une néguentropie structurale. Peut-être, mais cette néguentropie structurale est mise une fois pour toutes dans le crible, alors que le crible peut être utilisé de façon continue. De toute façon, donc, aussi grande que soit la néguentropie structurale, on pourra toujours rendre le bilan positif.)

D'autre part, M. Brillouin, pour résoudre le problème des copies que je ne peux aborder ici, est amené à faire une distinction très peu satisfaisante entre une information morte et une information vivante, la compensation n'intervenant, selon lui, que lors de la *lecture* de la copie, et non lors de son exécution.

Il me semble donc que la compensation doit être recherchée ailleurs, et j'ai essayé de montrer que toute mise en ordre d'un système, toute mise d'un système dans un état particulier exige une transformation d'énergie cinétique en chaleur, ou du moins d'énergie ordonnée en énergie désordonnée, ce qui représente une augmentation d'entropie. Autrement dit, *toute diminution de la variabilité quelque part dans un système, par exemple par une sélection, un tri ou autre mise en ordre, se paye par une augmentation de l'entropie-variabilité physique dans une autre* *partie de ce même système*. Ce principe est applicable tel quel à la compensation physique et je ne lui connais pas d'exception. Il n'est, en particulier, pas contredit par les phénomènes auxquels s'applique le théorème n° 7 de Shannon car là aussi, la diminution de variabilité est payée par une nécessaire augmentation de la variabilité physique en quelqu'endroit du traducteur. C'est donc sous cette forme que je verrais un principe de Carnot généralisé, applicable également à la théorie de l'information et à la physique.

Il ne me reste plus qu'à évoquer brièvement quelques problèmes que je n'ai pu aborder, mais qui ont un rapport assez étroit avec mon sujet.

Tout d'abord, il est clair que l'application des notions de variabilité et de spécificité ne se limite pas à la thermodynamique et à la théorie de l'information. On peut, par exemple, parler de la variété, de l'entropie du style d'un auteur : c'est une mesure de sa richesse et de sa diversité. On peut parler également de l'entropie d'une société, d'un groupe social : par exemple, une société libérale, individualiste, aura certainement une entropie plus grande qu'une société dirigiste et égalitaire. L'entropie peut même servir à caractériser la liberté laissée à l'individu ou du moins la liberté dont il fait usage.

La notion de *spécificité* est également utilisable dans la vie de tous les jours : certains objets, certaines structures n'ont une valeur pour nous que s'ils se trouvent dans un état particulier parmi tous les états qu'ils peuvent prendre. Ceci s'applique naturellement au message : un message n'a pour nous de la valeur que s'il se trouve dans un état bien précis, mais il en est de même pour un tableau, une mosaïque, une mélodie. Si l'on perturbe ces structures, elles perdent en même temps leur spécificité et leur valeur. Mais il ne faudrait pas en conclure que toute structure, par le fait même qu'elle est spécifique, acquiert de la valeur. L'originalité n'est pas une valeur en soi; c'est au contraire la valeur qui, du fait qu'elle est liée à des états bien particuliers, impose une certaine spécificité. Comme il faut se garder de confondre variabilité et spécificité, il faut se garder de confondre spécificité et valeur. Il faut d'ailleurs également être prudent dans l'identification de la spécificité avec la quantité d'information; on peut attribuer à un message une certaine spécificité sans qu'il ne transporte ou conserve aucune information.

Les structures spécifiques obéissent également à un principe de Carnot, le même que celui de la théorie de l'information, qui prescrit que la spécificité d'une structure abandonnée à elle-même ne peut que diminuer. Et ici encore se posent des problèmes au sujet de l'évolution biologique d'une part, de la création artistique, technique ou scientifique d'autre part, qui toutes deux augmentent la spécificité. Cette augmentation de spécificité est-elle de même nature que celle du choix ? Nous ne pouvons naturellement pas aborder ces problèmes ici et je renvoie ceux que cela pourrait intéresser à mon ouvrage *Information, thermodynamique, vie et pensée*, publié chez Gauthier-Villars.

Et, pour terminer, je vais résumer brièvement les points principaux de ma conférence.

J'ai tout d'abord exposé la théorie de l'information de façon superficielle, comme on le fait quelquefois, en faisant de l'entropie-variabilité elle-même une mesure de la quantité d'information. J'ai montré ensuite à quels paradoxes cette identification conduisait. On est donc amené à définir une autre grandeur, différence de deux entropies, la spécificité; cette spécificité est la véritable mesure de la quantité d'information. J'ai montré en passant comment on pouvait interpréter cette quantité d'information soit objectivement, soit subjectivement, objectif et subjectif ayant ici un sens bien particulier.

Dans une seconde partie, j'ai montré comment ces notions peuvent être appliquées à la thermodynamique et comment l'entropie, qui avait tout d'abord été interprétée par Boltzmann comme une spécificité négative, a reçu de Planck une autre interprétation, dite «statistique», et qui fait d'elle une variabilité.

Enfin, dans une dernière partie, j'ai parlé, en rapport avec ces notions, des principes de Carnot de la thermodynamique et de la théorie de l'information, ainsi que d'un éventuel principe de Carnot généralisé.