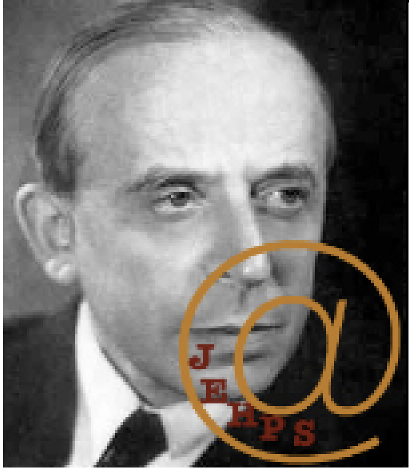

Journ@l Electronique d'Histoire des Probabilités et de la Statistique

Electronic Journ@l for History of Probability and Statistics

Vol 3, n°2; Décembre/December 2007

www.jehps.net

Information et entropie. Un double jeu avec les probabilités.

MATHIEU TRICLOT[1]

Résumé

Shannon et Wiener ont donné en 1948 une définition de la quantité d'information, en relation avec les problèmes de l'ingénierie des communications. La cybernétique s'est emparée de cette définition, en cherchant à penser la dimension physique de l'information au moyen d'une analogie entre information et entropie. L'article compare deux versions de cette analogie, chez Wiener et Brillouin, et propose de rapporter les divergences observées à l'interprétation duelle, épistémique ou fréquentiste, des probabilités mobilisées par la définition de l'information. Le concept de quantité d'information apparaît alors comme un mixte instable d'éléments épistémiques et fréquentistes.

Abstract

In 1948 Shannon and Wiener introduced a new measure of information in the field of communication engineering. In order to qualify the physical dimension of information, Cybernetics took over this measure and built an analogy between information and entropy. This article compares two versions of the analogy, in Wiener and Brillouin's work. It suggests that the divergence arises from the dual interpretation, belief or frequency-type, of the probabilities used in the information definition. The concept of information quantity thus appears as an unstable mix of belief-type and frequency-type elements of probability.

INTRODUCTION

L'objet de cet article est d'explorer les interprétations divergentes du concept mathématique d'information, tel qu'il a été introduit, au tournant des années 1940-1950, dans le sillage du mouvement cybernétique aux Etats-Unis. Nous avançons l'hypothèse selon laquelle ces divergences se rapportent, pour partie, à la manière d'interpréter l'usage des probabilités dans la définition de la quantité d'information.

On pourra alors considérer que les probabilités jouent un double jeu dans cette définition cybernétique de la quantité d'information. Au niveau technique, elles autorisent une formulation mathématique performante, qui englobe des questions échappant aux définitions non probabilistes antérieures. Mais les probabilités interviennent à un second niveau, philosophique plutôt que technique, dès lors que l'on s'interroge sur la signification du concept d'information ainsi produit et son application hors du strict champ de la théorie des

[1] Laboratoire *Récits* EA 3897, Université de Technologie Belfort-Montbéliard, 90010 Belfort, mathieu.triclot@gmail.com

télécommunications[2]. Il n'est pas impossible que les probabilités jouent ici un second rôle, à travers la question de l'interprétation « objective » ou « subjective » que l'on peut faire du formalisme. La clé de l'extension du concept d'information opérée par la cybernétique repose en effet sur une analogie entre l'information et l'entropie ; or, il apparaît que la signification de cette analogie peut elle-même être rapportée la façon dont on se représente les probabilités.

Ce double jeu des probabilités dans la définition de l'information nous entraîne sur un terrain glissant entre les sciences et la philosophie, en compagnie d'objets au statut épistémologique incertain. Dès lors qu'il existe une définition, ou même un ensemble de définitions, stable et bien accepté de l'information, que peut bien signifier l'existence d'une pluralité d'interprétations ? Quel est son statut ? Que faut-il entendre ici, au juste, par « interprétation » ? Comment peut-on hésiter sur le sens d'une formule, *a fortiori* aussi limpide que celle de la quantité d'information ? Mais, précisément, où réside « le sens » d'une telle formule ?

Cette situation épistémologique curieuse – une formule univoque pour des interprétations opposées – prend corps à l'intersection de trois ensembles de faits. Les travaux de Claude Shannon et Norbert Wiener, publiés de manière quasi simultanée en 1948, introduisent tout d'abord des définitions mathématiques de la quantité d'information, reconnues équivalentes[3]. Schématiquement, on dira que la définition de Shannon est construite à partir du cas discret (pour une suite de symboles), celle de Wiener à partir du cas continu (la variation d'un signal). Ces définitions ne sont pas absolument nouvelles, au sens où elles reprennent, de manière explicite, des éléments existants dans la tradition des ingénieurs en télécommunications, mais elles se distinguent par leur insertion dans un cadre conceptuel unifié, qui autorise en pratique la résolution d'une large classe de problèmes, du codage au traitement du signal, de la conception des filtres à celle des amplificateurs, autrefois considérés comme disjoints. Cette extension de la définition et la nouvelle synthèse qu'elle autorise résultent de l'introduction d'un formalisme de type probabiliste. Nous tenons notre fait n°1 : la définition mathématique classique de la quantité d'information, celle proposée par Wiener et Shannon en 1948, qui s'impose par la synthèse qu'elle opère au sein des disciplines de l'ingénieur, repose sur une innovation conceptuelle majeure, le recours aux probabilités. Sans probabilités, il n'y a pas de définition de Shannon-Wiener. Les définitions antérieures ne faisaient pas usage du calcul des probabilités.

---

[2] Il faut distinguer le problème de « la dimension signifiante de l'information » de ce que nous appelons ici « la signification du concept d'information ». Dans le premier cas, il s'agit du fait que la mesure de l'information ne prend pas en compte la signification des messages. Dans le second, il s'agit de la portée que l'on doit accorder à la définition mathématique de l'information, hors du domaine des télécommunications. Cette question nous renvoie au problème de la dimension physique de l'information, plus qu'à celui de la dimension signifiante des messages.

[3] [Shannon, 1949] et [Wiener, 1948] sont les textes canoniques, qui portent à la connaissance du public des travaux datant de la période de guerre. [Shannon, 1949] reprend sous forme de livre, augmenté d'une préface de vulgarisation par Warren Weaver, les articles parus en 1948 dans le *Bell System Technical Journal*. Ces articles se situent dans la continuité du rapport de Shannon consacré pendant la guerre aux questions de cryptographie (sur le sujet, on consultera [Segal, 2003]). [Wiener, 1948] puise largement quant à lui dans le rapport de 1942 au NDRC, « Extrapolation, Interpolation, and Smoothing of Stationary Time Series », finalement publié en 1949, sous le titre plus ramassé de *Time Series* [Wiener, 1949]. L'équivalence entre les définitions est reconnue par Wiener et Shannon eux-mêmes. Wiener cite Shannon pour avoir « développé [simultanément] une théorie statistique de la quantité d'information » ([Wiener, 1948], p. 10). Shannon cite Wiener avoir formulé le premier, de manière explicite, la théorie de la communication sous une forme statistique ([Shannon, 1949], p. 85]). Cela ne règle pas la question, au-delà du cadre américain, du rapport entre les définitions de Wiener et Shannon et celles produites dans d'autres contextes, comme celles de Fisher ou Gabor ([Segal, 2003], pp. 294-304).

L'histoire pourrait s'arrêter là – et la construction de cette définition de la quantité d'information est déjà une histoire en soi –, mais il se trouve que, pour toute une série de raisons qui dépassent le cadre de cet article, cette même définition de la quantité d'information a été instituée comme emblème d'un discours scientifique plus vaste, aux ambitions nettement plus dévorantes, qui loin de se cantonner aux problèmes circonscrits de la transmission d'information dans le cadre de la théorie des télécommunications, a au contraire prétendu dresser des ponts entre la physique statistique, les sciences de l'ingénieur, les sciences du vivant, les sciences de la cognition et les sciences sociales ; en bref, rien moins que redessiner tout l'édifice des savoirs à partir du concept d'information. Ce mouvement, que Wiener a baptisé du nom de cybernétique, s'est défini comme science générale du contrôle et de la communication, chez l'animal et la machine. C'est au cours de ce mouvement d'extension de la notion d'information, cette fois-ci hors du domaine des télécommunications, qu'apparaissent un certain nombre de clivages quant à la signification de l'information. Il s'agit alors, au-delà des problèmes de la construction du code le plus efficient ou du filtre le moins destructeur, de s'interroger sur ce que information veut dire en réalité, pour transférer le concept d'un champ à un autre. Peut-on parler d'information à propos d'un organisme ou d'une machine, d'un cerveau ou d'un calculateur, comme l'on a appris à le faire, au cours de la guerre, pour désigner les opérations de machines aussi différentes qu'un télégraphe, un téléphone, un radar, un calculateur ou une batterie de dca ? Ce deuxième fait nous conduit à l'interface entre un formalisme mathématique local, celui de Shannon et Wiener, et son interprétation plus large, philosophique, qui s'énonce comme telle dans des chapitres ou des revues de philosophie des sciences : de quoi s'agit-il lorsque l'on nous parlons d'information[4] ?

Fait n°3. Ces tentatives d'interprétation s'incarnent, pour les années 1950, et au-delà, dans la question de la signification physique de l'information. Se demander qu'est-ce que c'est que de l'information revient alors à se demander quelle quantité physique est l'information. Etre quelque chose en réalité et être une quantité physique sont ici naturellement considérés comme synonymes. Si l'information n'est selon le mot de Wiener « ni matière ni énergie », elle n'est certainement pas pour la cybernétique un objet idéal ou hors nature[5] ; bien au contraire, la cybernétique s'est toujours conçue comme une région au sein du vaste empire de la physique ; une « terra incognita », un « no man's land » ouvert à l'ambition des pionniers certes, mais une région tout de même[6].

Le problème de la signification physique de l'information se cristallise en cybernétique autour de l'analogie entre l'information et cette quantité physique particulière qu'est l'entropie ; analogie conquise dans le cadre de l'expérience de pensée du démon de Maxwell. Avec les cybernéticiens, on sera donc amené à dire que l'information désigne *en réalité* l'inverse de l'entropie. Or, sur cette question, les tenants de l'analogie information-entropie dans les années 1950 ne présentent pas un front uni, et il existe, au contraire, des divergences majeures entre l'interprétation de cette analogie chez Wiener et la version canonique qu'en a donné Léon Brillouin avec son célèbre principe de néguentropie de l'information[7]. Ce clivage

---

[4] A titre d'exemple, « Purpose, Behavior and Teleology », l'article fondateur du mouvement cybernétique, en 1943, paraît dans la revue *Philosophy of Science* [Rosenblueth, Wiener, Bigelow, 1943]. Il ne s'agit pas là d'un cas isolé (que l'on songe aux articles de Turing). Ce type de discussion « philosophique » apparaît constamment dans le sillage de la cybernétique. Il est possible de soutenir qu'il y joue un rôle effectif.

[5] [Wiener, 1948], p. 132.

[6] [Wiener, 1948], p. 2.

[7] Du côté des critiques de l'analogie entre information et entropie – que l'on songe par exemple aux critiques de Popper, sur un plan épistémologique, ou à celles de Bennett, adossées à la physique du calcul –, la distinction entre l'analogie portée par Wiener et celle portée par Brillouin n'apparaît pas ([Popper, 1957], [Popper, 1974], [Bennett, 1982]). En réalité, c'est toujours la version de Brillouin, la plus développée, qui est visée. Or, la

nous renvoie à la composante fondamentale du projet cybernétique et au rôle que celle-ci a toujours accordé à la physique statistique dans la construction de son programme de recherches.

Notre objet d'étude prend forme. Nous avons une définition mathématique unifiée, qui se présente à nous auréolée d'indéniables succès théoriques et pratiques, mais aussi plusieurs manières de se représenter ce que peut signifier en réalité cette définition mathématique. Cette question de « la signification en réalité » apparaît au cours de l'opération épistémologique risquée qui consiste à passer d'un domaine de savoir à un autre, des sciences de l'ingénieur à la physique, de la physique au vivant, opération qui requiert pour son propre compte une représentation commune, transversale, de ce que peut être l'information. Cependant, au lieu de produire la représentation univoque et universellement transférable originellement escomptée, la cybernétique accouche d'une hydre épistémologique, d'un concept d'information à plusieurs têtes dont aucune ne consent à tourner dans le même sens. Comment trancher entre interprétations divergentes ? Que faire de ces mixtes franchement inconvenants de mathématiques, de physique et de philosophie ?

Les divergences d'interprétation des probabilités – un formalisme commun pour des représentations irréconciliables, au moins dans leurs versions dogmatiques, de ce que la probabilité signifie – offrent un modèle séduisant pour commencer à penser la situation du concept d'information produit par la cybernétique. Voici qu'un même ensemble de calcul peut être conçu pour représenter soit les règles régissant les degrés de croyance (interprétation épistémique) soit le comportement à long terme des fréquences relatives dans une suite (interprétation fréquentiste)[8]. Ce modèle, qui pointe un écart entre un formalisme et sa représentation, est séduisant parce qu'il nous conduit à ne pas assimiler trop vite questions de représentations et questions d'idéologie. Les historiens de la période cybernétique ont parfois été tentés, suivant en cela le discours des protagonistes eux-mêmes, par une partition stricte entre les bonnes mathématiques pour ingénieurs et le mauvais supplément idéologique des analogies philosophiques[9]. On peut penser que la situation est en réalité un peu plus

version de Wiener, par son interprétation objective des probabilités comme par son insistance sur la dimension physique de l'information, n'apparaît finalement pas si éloignée des critiques que Popper ou Bennett adressent à Brillouin.

[8] J'emprunte ici terminologie et définitions à [Hacking, 2001], pp. 145-146. Il est possible que cette distinction ne possède pas la portée historique que Hacking lui attribue ([Desrosières, 2006], p. 3 ou [Martin, 2006], p. 18). Une telle question dépasse cependant de loin le cadre de cet article, qui se contente de proposer une « application » de cette distinction à un cas particulier.

[9] Pierce, acteur et historien de la théorie des communications, note avec ironie combien « le terme cybernétique est utile pour ajouter un peu de glamour à quelqu'un, à un sujet ou même un livre », avec l'espoir que « sa présence ici ajoutera un peu de glamour à ce livre-ci » ([Pierce, 1961], p. 228). Il faut dire que Shannon lui-même, dans son éditorial de 1956 aux *IRE Transactions in Information Theory*, avait déjà dénoncé « la fanfare cybernétique » ([Shannon, 1956]), invitant à un recentrage sur le noyau dur mathématique de la théorie. [Segal, 2003] distingue, de son côté, entre le « concept » (mathématique) d'information et la « notion » d'information, plus floue, circulant entre les différents domaines, sous le mode de la métaphore, de l'analogie ou du modèle. La notion d'information se présente sous une forme ambivalente « tout à la fois heuristique et potentiellement fallacieuse ». La situation des probabilités nous invite à considérer les débats autour de la signification de l'information sous un jour un peu différent, en insistant sur l'existence de ce que Hacking a appelé des « points de blocage » qui nous renvoient à des controverses philosophiques, incessamment reprises sous le vêtement du jour, mais qui ne se laissent pas aisément résorber ([Hacking, 1999], pp. 91-92). Il ne s'agit donc pas ici uniquement de la dimension heuristique et prospective des concepts, encore moins des mauvais usages, mais de philosophie entendue comme discussion rationnelle, entreprise de clarification, au point où nos schèmes conceptuels nous mènent à des antinomies ; ici, le statut des probabilités en physique ou le rapport entre matière et signification. Se pose la question de savoir si on peut repérer une dimension philosophique qui soit immanente à l'activité scientifique et technique, dès lors qu'elle est portée par les acteurs eux-mêmes, et non seulement ajout externe, ainsi que le suggèrent d'autres études, sur d'autres sujets, comme, par exemple [Hacking, 1975] ou [Paty, 1993]. Il va de soi que la présente étude ne constitue pas un travail d'histoire des sciences au sens strict,

compliquée ; complication que la situation des probabilités permet d'aborder. Celle-ci nous montre en effet, premièrement, que les questions d'interprétation philosophique ne sont pas nécessairement des questions triviales ou oiseuses, et deuxièmement, qu'elles interviennent dans le développement du calcul lui-même, soit sous la forme de questions sur les fondements, soit en privilégiant tel ou tel axe de recherche.

Dans le cas de l'information, on ne peut pas vraiment dire que les interprétations interviennent dans le développement mathématique de la notion, qui sous sa forme cybernétique est largement clos en 1948. En revanche, elles interviennent manifestement dans les développements ultérieurs de la théorie algorithmique, tout comme elles jouent un rôle clé dans la définition du programme cybernétique. Ajoutons que ces conflits de représentation sont loin d'être triviaux et qu'ils réinvestissent certainement, à l'image du problème de l'induction pour la théorie des probabilités, de vieux problèmes laissés en friche par la philosophie, comme celui du rapport entre la matière et la forme ou de l'inscription matérielle du signe. En bref, le modèle du conflit des probabilités nous conduit à prendre au sérieux cette zone de discours marquée au sceau de l'incertitude entre le calcul et la philosophie, plutôt que de la rejeter dans l'enfer des pensées mal tournées. On y gagne la possibilité de considérer l'immense production philosophique qui accompagne le développement de la cybernétique autrement que comme un développement malheureux, uniquement explicable par la soif de gloire de quelques savants en mal de reconnaissance. Ce n'est pas rien. Prenons au sérieux les mixtes de mathématique et de philosophie quand ils se présentent, et considérons leur situation stratégique dans la discussion des programmes de recherche. Il s'agit de faire droit à une raison spéculative exercée à propos de la raison démonstrative.

Mais, il se trouve aussi que, dans le cas présent, la situation du calcul des probabilités nous offre un peu plus qu'un secours épistémologique général pour prendre au sérieux les conflits philosophiques autour d'un formalisme. Il est en effet possible de rapporter pour une part le conflit autour des représentations de l'information à la manière d'interpréter l'usage des probabilités dans la définition de l'information. Le concept d'information présente cette propriété intéressante de se situer juste sur la ligne de tension entre interprétation fréquentiste et interprétation épistémique. La définition probabiliste de l'information ne fait aucunement appel à une quelconque mesure des degrés de certitude, mais se construit uniquement par l'étude des sources de messages assimilées à des générateurs aléatoires. Mais, définie ainsi de manière intégralement stochastique, la quantité d'information s'interprète pourtant spontanément de manière épistémique, comme une mesure du degré de connaissance. Qu'est-ce que de l'information, sinon une mesure de la « dispersion » des probabilités gouvernant le comportement d'un générateur aléatoire ? Qu'est-ce que de l'information, sinon une mesure de la façon dont notre connaissance peut être augmentée à la réception d'un message ? Il n'est pas impossible que cette double approche de l'information, inextricablement épistémique et fréquentiste, ait joué son rôle dans la constitution de représentations opposées de ce que désigne l'information, en particulier dans le cas des différences entre Wiener et Brillouin.

Nous avancerons cette hypothèse, en examinant d'abord le rôle que l'introduction des probabilités joue dans la définition de l'information chez Shannon et Wiener, puis les éléments qui autorisent la construction d'une analogie au plan physique entre information et entropie, avant de décrire enfin les divergences sur la manière d'interpréter cette analogie, et ce qu'elles doivent à l'interprétation de la probabilité.

---

mais une réflexion, « interne », sur les paramètres, d'ordre épistémologique, qui rendent possible un partage entre deux interprétations de l'analogie information-entropie.

1. UNE DEFINITION PROBABILISTE DE L'INFORMATION

L'histoire de la constitution du concept d'information au sein de la théorie des télécommunications est bien connue. On pourra se rapporter sur le sujet à [Mindell, 2002], [Pierce, 1961], [Segal, 2003]. Shannon ouvre d'ailleurs « A Mathematical Theory of Communication » par la mention des travaux de ses prédécesseurs aux Bell Labs, Harry Nyquist et Ralph Hartley, qui proposaient dès les années 1920 les premières pistes pour une quantification de l'information.

En 1924, dans « Certain Factors Affecting Telegraph Speed » Nyquist a introduit une mesure élémentaire de la « vitesse de transmission de l'information »[10]. Nous disposons de deux possibilités pour accroître la vitesse de transmission d'une ligne de télégraphe : soit nous augmentons directement la vitesse de transmission, ce qui pose des problèmes drastiques de fiabilité du fait de l'atténuation du signal sur les longues distances, soit nous augmentons le nombre de messages transmis en même temps, en jouant sur les possibilités de codage offertes par les différentes valeurs de courant. Nyquist propose alors naturellement de mesurer la vitesse de transmission de l'information comme une fonction du nombre de valeurs de courant disponibles en télégraphie. Si m est le nombre de valeurs de courant, k une constante exprimant la vitesse de transmission de la ligne, alors la vitesse de transmission de l'information en valeurs de courant par seconde (W) est simplement :

$$W = k \log_2 m$$

Nyquist construit ensuite une estimation du rapport entre le nombre de valeurs de courants et la bande passante – première approche du problème de l'échantillonnage –, ce qui lui permet au final d'exprimer la relation entre vitesse de ligne, bande passante et valeurs de courant.

Ce résultat, qui peut bien nous apparaître trivial au vu des réalisations ultérieures, est cependant représentatif des changements qui interviennent dans la manière de considérer les lignes de télécommunications à partir des années 1920, dès lors que s'ajoute à la problématique de la transmission de puissance un ensemble de questions portant sur les possibilités de codage ou la mise en forme des signaux. Voici que nous apprenons peu à peu à interroger la performance des lignes non plus comme une simple affaire de transfert électrique, mais comme la transmission d'une réalité spécifique, le message, qui possède ses propres lois. Outre l'échelle logarithmique, que l'on retrouvera dans toutes les définitions ultérieures et qui s'impose pour maintenir l'additivité des éléments d'information face au nombre d'éléments disponibles pour construire le message, l'article de Nyquist introduit ainsi de manière incomplète nombre d'éléments qui seront décisifs pour l'ingénieur en télécommunications : le principe d'une construction des messages à partir d'un choix entre symboles binaires, la relation entre vitesse de ligne et bande passante, une première ébauche du principe d'échantillonnage ou encore une esquisse du concept de redondance. Le référent technique étroit de la télégraphie à courant continu empêche cependant d'aller beaucoup plus loin. La formule de Nyquist reste cantonnée à un choix entre symboles équiprobables et de durée équivalente. On ne trouve chez Nyquist aucune trace des théorèmes de codage fondés sur l'étude de la structure statistique de la source qui seront l'essentiel du travail de Shannon, pas plus que la prise en compte du bruit comme composante à part entière du problème des télécommunications.

---

[10] Nyquist n'emploie pas le terme « information », mais « intelligence », conformément à l'usage de l'époque. Voir, par exemple, « Electric Communication, Past, Present and Future », le discours de Frank Jewett, directeur des laboratoires Bell, à l'Académie des Sciences en 1935, qui se propose d'identifier sous le terme d'« intelligence » l'objet commun de tous les processus de télécommunications et de la science qui s'y applique. Cité par [Mindell, 2002], p. 136.

« Transmission of Information » de Ralph Hartley en 1928, l'autre article cité par Shannon, n'innove pas sur ce point. Hartley y généralise les résultats de Nyquist pour introduire la première mesure explicite de la quantité d'information. L'opération de communication est conçue comme un processus dans lequel l'émetteur sélectionne parmi un jeu de symboles une séquence qui forme le message. La quantité d'information mesure ainsi la quantité de choix impliqués dans un processus de sélection parmi un ensemble de symboles. Plus il y a de symboles possibles, plus il y a de choix et plus il y a d'information. Soit n le nombre de symboles sélectionnés et s le nombre de symboles disponibles, la quantité d'information (H, pour Hartley) s'exprime alors simplement :

$$H = n \log s$$

Ainsi, si nous disposons de deux symboles élémentaires (0 et 1) et que nous choisissons de transmettre un message formé d'un seul symbole, la quantité d'information vaut 1. L'unité d'information représente le choix élémentaire entre deux symboles équiprobables et indépendants. Le jet d'une pièce de monnaie équilibrée est un exemple d'une source produisant 1 bit d'information par lancer.

A la différence de Nyquist, la formule de Hartley est donc faite pour s'appliquer à n'importe quel système de transmission, le but avéré étant de fournir une unité de mesure universelle. Cependant, pas plus que Nyquist, Hartley n'est en mesure de poser rigoureusement le problème de l'efficience du codage, qui suppose de s'intéresser à la structure statistique de la source et de dépasser le modèle du choix entre symboles équiprobables, ou d'intégrer mathématiquement la question du bruit de ligne. L'objet information prend forme chez Nyquist et Hartley, au sens où émerge progressivement une dimension de l'activité de télécommunication qui n'est pas réductible au simple aspect énergétique de la transmission. Mais les éléments fondamentaux de ce qui sera la théorie de l'information échappent encore à Nyquist ou Hartley.

Il est désormais facile de repérer ce qui distingue les travaux de Shannon et Wiener des travaux des années 1920 et 1930. Shannon et Wiener innovent non en proposant une définition mathématique de l'information – celle-ci existe déjà chez Hartley –, mais en fournissant une solution rigoureuse au problème du codage des messages (Shannon) et à celui du filtrage et de la prédiction des signaux (Wiener). Ces travaux impliquent en retour une extension (et une sophistication) considérable de la définition de la quantité d'information ; extension rendue possible par le recours au formalisme probabilitaire.

L'étude du canal discret et sans bruit, dans le premier des deux articles de Shannon en 1948, mathématiquement la plus simple, offre un bon échantillon de ce que l'introduction des probabilités apporte à la théorie de l'information. Shannon prend pour point de départ le principe posé par Hartley : la quantité d'information doit être fonction de la quantité de choix opérés par l'émetteur dans la construction du message.

« Supposons que nous ayons un ensemble d'évènements possibles, dont les probabilités d'occurrence sont $p_1, p_2, \ldots, p_n$. Les probabilités sont connues, mais c'est tout ce que nous connaissons de l'évènement à venir. Pouvons-nous trouver une mesure du nombre de choix impliqué dans la sélection de l'évènement ou une mesure de l'incertitude du résultat ?[11] »

Cette description permet de retrouver l'évaluation spontanée de la quantité d'information : plus il y a de messages disponibles entre lesquels choisir, plus l'incertitude pesant sur le message choisi du point de vue du destinataire sera élevée, et plus l'information acquise à la réception du message sera importante. Autrement dit, il y a plus d'information dans un message hautement improbable que dans un message dont le contenu est entièrement

[11] [Shannon, 1949], p. 49.

prévisible. A la limite, si je n'ai qu'un seul message à transmettre, alors il n'existe aucune indétermination à la réception, la quantité d'information est nulle et l'opération de communication elle-même superflue. Reste à trouver une expression mathématique pour exprimer cette quantité d'indétermination, en fonction de ce que l'on connaît de la structure statistique de la source ; une expression qui soit à la fois mathématiquement consistante, conforme à nos intuitions de base sur les variations de la quantité d'information et féconde du point de vue des applications. Shannon généralise alors les définitions de Nyquist et Hartley en retenant la somme du logarithme des probabilités d'apparition des différents symboles comme expression de la quantité d'information :

$$H = - k \Sigma p_i \log_2 p_i$$

K est une constante due au choix d'une unité de mesure. Le signe « moins » permet de rendre le résultat positif, conformément à l'intuition que l'information est une quantité positive. Le logarithme maintient l'additivité des quantités d'information. Le choix de la base 2 pour le logarithme impose le digit binaire, ou bit, comme unité d'information, correspondant au choix élémentaire entre deux messages (0 ou 1) également probables.

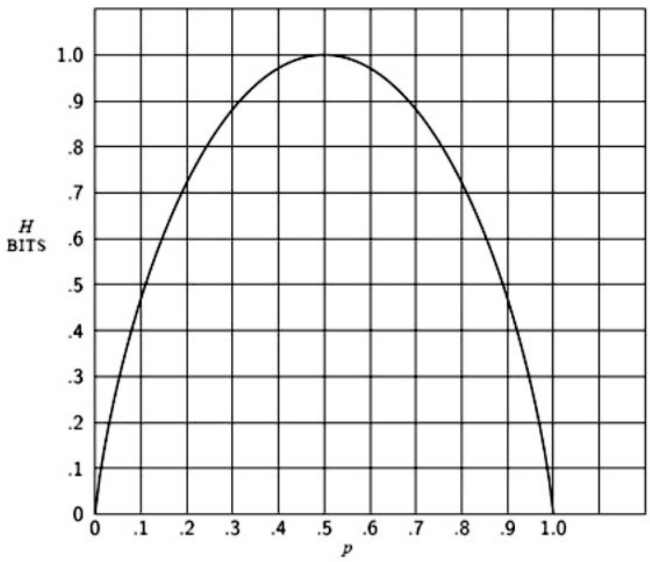

Fig. 1 – Valeurs de H pour deux symboles. [Shannon, 1949], p. 50.

On peut vérifier que H augmente lorsque les probabilités convergent (situation d'incertitude maximale) et diminue lorsqu'elles divergent (il y a moins de messages probables). Si pour $p_0 = p_1$, $H = 1$, on obtient par exemple $H = 0{,}811$ bit pour $p_1=3/4$, $p_0=1/4$. Enfin si p_1 ou p_0 vaut 1, alors H vaut bien 0.

L'intérêt de cette extension probabiliste de la définition de la quantité d'information est immédiat, non comme mesure de l'information en soi, mais comme expression d'une limite idéale pour nos pratiques de codage. A quoi sert la quantité de Shannon ? Elle exprime le nombre de symboles binaires requis pour encoder au mieux, avec le minimum de pertes, les messages produits par une source. Tout l'enjeu du travail de Shannon consiste ainsi à nous permettre d'économiser sur la capacité du canal (ou le temps de transmission), en profitant de nos connaissances sur les structures statistiques des sources de message.

Prenons un exemple afin d'illustrer la fonction de la quantité H. Shannon montre que le nombre de messages probables produits par une source est fonction de H. Si H est la quantité d'information de la source, M le nombre de messages qu'elle produit, alors pourvu que l'on prenne des messages suffisamment longs, on peut considérer qu'il n'y a que 2^{MH} messages probables. La probabilité des messages restants (par exemple, une longue suite d'un même caractère indéfiniment répété) est suffisamment faible pour être négligée. On peut donc envisager de ranger les messages du plus probable au moins probable, jusqu'au $2^{MH\text{-ième}}$ message, en leur attribuant à chacun un numéro (de 1 à 2^{MH}), en étant quasiment certain d'avoir un numéro pour n'importe quel message produit par la source. Si nous prenons une source qui fonctionne avec deux symboles équiprobables et produit des messages de 1000 caractères, MH vaut alors 1000. D'après le résultat de Shannon, il y a donc 2^{1000} messages probables différents, soit exactement le nombre de messages de 1000 caractères que l'on peut écrire avec deux symboles. Pour encoder un message de 1000 caractères produit par cette source dont l'entropie est maximale, nous aurons toujours besoin d'au moins 1000 digits binaires. Aucune économie n'est possible ici. Mais supposons maintenant que notre connaissance de la structure statistique de la source s'améliore et que nous apprenions que les probabilités d'apparition des symboles ne sont plus égales, mais que $p_0= ¾$ et $p_1= ¼$. En conservant nos messages de 1000 caractères, MH vaut maintenant 811. Ainsi, alors qu'il y a 2^{1000} messages possibles, il n'y a plus que 2^{811} messages probables. Nous pouvons donc associer chacune des combinaisons des 811 digits binaires à un message probable de 1000 caractères. A condition d'encoder par bloc (ici de 1000 caractères), 811 caractères nous suffisent à écrire tous les messages probables de 1000 caractères.

La quantité d'information ne désigne donc rien d'autre que le nombre de digits binaires requis pour encoder un message, dès lors que l'on choisit des blocs suffisamment longs. Il s'agit bien entendu d'une limite théorique. Non seulement, on n'est jamais sûr d'atteindre la compression optimale – la théorie de Shannon ne comporte pas (encore) de méthode constructive pour obtenir de manière systématique H –, mais la limite théorique n'est pas toujours souhaitable en pratique, où l'on souhaite pouvoir conserver au message une certaine redondance pour lutter contre le bruit. En pratique, coder revient à échanger une redondance naturelle, celle de la source, contre une redondance choisie, celle du code, de façon à optimiser la résistance aux fluctuations aléatoires du bruit de ligne.

Si les avantages de cette définition probabiliste de l'information sont évidents, en particulier pour la formulation du problème du codage qui voit l'ingénieur mettre à son service la connaissance de la structure statistique des sources, le recours aux probabilités introduit cependant plusieurs limitations drastiques quant à l'usage qui peut être fait de la quantité d'information.

Les commentateurs ont souvent insisté, à commencer par le premier d'entre eux, Warren Weaver dans sa préface à la réédition sous forme de livre des articles de Shannon, sur le fait que la définition de Shannon laisse totalement de côté le fait de savoir si l'information transmise est ou non signifiante[12]. Information et signification sont du point de vue de la théorie de Shannon deux concepts disjoints. Un message signifiant ne possède pas nécessairement une quantité d'information plus grande qu'un message dénué de sens. La problématique de l'ingénieur en télécommunications exigerait donc de mettre totalement de côté le contenu des messages pour ne plus s'intéresser qu'à leur forme, fidèle en cela à l'éthique de l'employé des postes qui s'abstient de lire le courrier qui lui passe entre les mains.

Mais cette limitation, si souvent mise en avant, n'est rien comparée aux contraintes que fait peser une définition probabiliste de l'information. Ajoutons que telle qu'elle est souvent exprimée – on ne peut distinguer entre un message signifiant et un message qui ne l'est pas –

[12] [Shannon, 1949], p. 8.

elle est tout bonnement inexacte. Non que l'on sache distinguer entre un message signifiant et un message qui ne l'est pas, mais parce que l'on ne peut chez Shannon aucunement distinguer des messages individuels du point de vue de leur quantité d'information. La quantité d'information est toujours chez Shannon la quantité d'information d'une source de messages, définie par son ensemble de probabilités, et non la quantité d'information d'un message individuel. Un message ne possède pas de quantité d'information si ce n'est celle qu'il partage avec l'ensemble des autres messages produits par la même source. Cette limitation est plus fondamentale que le soi-disant oubli de la signification[13]. En particulier elle limite considérablement les possibilités d'application du calcul dès lors que l'on s'éloigne du schéma des télécommunications et que l'on se met en tête de considérer la quantité d'information contenue, par exemple, dans un brin d'ADN, un programme ou un organisme, pour ne prendre que quelques uns des fétiches cybernétiques les plus en vue. Il semble que l'application de la théorie de Shannon, en particulier dans le champ de la biologie, ne puisse conduire qu'à une déception éternellement recommencée.

L'introduction d'une définition probabiliste de l'information chez Shannon constitue donc une force et une faiblesse. Une force en ce qu'elle autorise la position rigoureuse du problème du codage des messages. Une force en ce qu'elle permet de traiter élégamment le bruit de ligne comme une source additionnelle qui se compose avec la source de messages légitime, injectant de l'aléatoire dans les messages transmis. Une faiblesse parce qu'elle réduit les calculs de la quantité d'information à la source des messages et interdit ainsi, sauf contorsions, les applications immédiates hors du champ de l'ingénierie. Or, jamais personne n'a semblé près à considérer que l'information se réduisait à l'information échangée dans le processus de télécommunications tel qu'il a été modélisé par Shannon. De là, la tentation d'étendre la définition à d'autres contextes. Une telle opération n'est pas nécessairement illégitime, mais elle nous entraîne à coup sûr dans une zone de turbulences spéculatives, nous obligeant à renégocier ce que peut signifier, hors théorie des télécommunications, la formule de la quantité d'information. Il se trouve que l'article de Shannon possédait pour cela, comme nous le verrons, un levier gigantesque dont le premier Archimède de passage ne pourrait s'empêcher de s'emparer pour soulever et remettre sur sa base le monde de l'information.

Il n'existe pas, à notre connaissance, de comparaison systématique entre le travail de Shannon et celui de Wiener en théorie de la communication. La reconnaissance mutuelle de l'équivalence des définitions, en dépit de quelques commentaires acerbes de part et d'autre sur les questions de priorité, a sans doute contribué à laisser la question hors du champ de l'étude historique. La technicité des résultats de Wiener, comparée à la lumineuse simplicité de l'exposé de Shannon, n'est sans doute pas pour rien non plus dans l'absence d'une telle comparaison. Pourtant si les résultats sont bien équivalents, ils ne sont pas exactement identiques ; la mesure de Wiener autorise par exemple des valeurs négatives. Surtout, ils relèvent de démarches mathématiques sensiblement différentes. La comparaison des travaux de Wiener et Shannon pourrait sans doute fournir un bon échantillon d'une étude des styles de raisonnement mathématique, dont on ne peut ici que tenter l'ébauche.

Le travail de Wiener se présente explicitement sous la forme d'une synthèse entre théorie statistique et ingénierie des communications. Wiener conduit une exploration systématique de ces deux domaines dont « l'unité méthodologique naturelle » avait jusqu'ici échappé à l'examen. Cette unité repose rien moins que sur l'existence d'un objet en commun : les séries temporelles[14].

---

[13] Ce point est bien mis en valeur, par comparaison avec la théorie de l'information algorithmique, dans [Delahaye, 1999], p. 20.
[14] [Wiener, 1949], p. 1.

Cette approche de l'information, plus englobante que celle de Shannon, implique de ne pas considérer la théorie de la communication comme une unité close. Deux différences sautent ainsi aux yeux, en première approche, dans la manière de constituer le domaine de l'information. Premièrement, la convergence entre théorie des probabilités et théorie de la communication est, chez Wiener, à double entrée ; il ne s'agit plus seulement de dire comme chez Shannon que la théorie de la communication a à être constituée comme une théorie probabiliste, en appliquant un formalisme existant aux problèmes du codage, mais aussi que la théorie statistique a à apprendre des outils mathématiques développés par les ingénieurs, en particulier l'usage de la transformée de Fourier. Wiener vise ainsi une fusion entre les deux domaines et non seulement une application des théories statistiques aux problèmes de l'ingénieur. Cette perspective apparaît liée au primat donné à l'analyse des ensembles de fonctions continues. Il est intéressant de noter que l'intuition de Wiener – la théorie des probabilités a à apprendre de la théorie de l'information – anticipe sur le renversement des années 1960 qui verront la nouvelle théorie de l'information algorithmique être mobilisée pour fournir un soutien inattendu au problème du fondement des probabilités, via la définition mathématique de la notion de suite aléatoire[15].

Deuxièmement, dans la continuité de ce rapport à double entrée entre théorie statistique et théorie de la communication pour l'analyse des séries temporelles, la théorie des probabilités est introduite d'une toute autre manière chez Wiener que chez Shannon, puisque Wiener opère systématiquement le lien entre théorie des probabilités et théorie physique. La théorie des probabilités ne se présente pas comme un ensemble clos ou un simple calcul détaché de ses applications. La source des résultats de Wiener est en effet à rechercher du côté de ses travaux pionniers sur la théorie probabiliste de la mesure et leur application à la modélisation des trajectoires dans un mouvement brownien, qui sont constamment évoqués dans les travaux des années 1940[16]. Le travail de Wiener conduit à situer d'emblée le traitement du signal dans un univers plus vaste, celui de la physique statistique et de son hypothèse ergodique, celui des phénomènes de fluctuations aléatoires comme le mouvement brownien. Information et probabilité s'ancrent naturellement dans le champ de la physique. Il y a de l'information (et de la probabilité) parce que nous habitons un univers irréversible et intrinsèquement marqué par les phénomènes aléatoires[17].

2. « L'INFORMATION EST L'INVERSE DE L'ENTROPIE »

Nous avons décrit dans ses grandes lignes la constitution de la notion scientifique d'information, en insistant sur le rôle joué dans les définitions de Shannon et Wiener par le recours au calcul des probabilités, principale innovation des définitions de l'année 1948. Mais

---

[15] La théorie algorithmique de l'information reprend le problème de la définition mathématique d'une suite aléatoire, posé mais non résolu par la théorie des collectifs de Von Mises. Il ne s'agit donc plus seulement de construire une mesure probabiliste de l'information, mais de reconstruire les fondements de la théorie (fréquentiste) des probabilités à partir d'un nouveau concept d'information algorithmique ([Von Plato, 1994], pp. 233-237, [Delahaye, 1999], pp. 29-64). Intuitivement, une suite finie aléatoire se définit alors comme une suite qui pour être calculée demande un programme aussi long qu'elle-même. La position de Wiener qui insiste sur la théorie de la communication comme un élément susceptible de féconder la théorie des probabilités, et non comme une simple application, sans dégager cependant la question de la définition des suites aléatoires, présente une analogie avec le renversement des années 1960.

[16] [Heims, 1980], [Masani, 1990], [Wiener, 1949] ou encore [Wiener, 1956] comportent des exposés accessibles de ces travaux.

[17] Voir, sur ce point la fable du démon, qui, vivant dans un temps inversé par rapport au notre, ne peut communiquer avec nous, dans le premier chapitre de *Cybernetics*. [Wiener, 1948], pp. 34-35.

la cybernétique ne s'est pas contentée d'une formule, elle a aussi voulu savoir ce que l'information voulait dire en réalité, question motivée en grande partie par la construction de son programme interdisciplinaire. Cette recherche de ce que l'information signifie, au-delà de la théorie des télécommunications, instaure un aller-retour curieux entre la définition mathématique et la notion intuitive. Loin de considérer la définition mathématique comme une manière de circonscrire une fois pour toutes la notion intuitive, la cybernétique s'est au contraire servie de la définition mathématique comme d'un tremplin pour atteindre en l'information une réalité physique dernière. Le développement de ce discours visant à construire une interprétation physicaliste de l'information constitue l'un des traits les plus marquants de la cybernétique, qui la distingue à coup sûr des approches majoritaires de l'information réduite à un pur agencement de symboles.

Les cybernéticiens disposaient pour construire une telle interprétation d'un atout majeur, qui ne pouvait qu'entrer en résonance avec la manière dont Wiener avait situé la question de l'information. La définition de la quantité d'information proposée par Shannon présente en effet une analogie frappante avec la définition d'une quantité physique particulière, l'entropie. Cette similitude entre les fonctions conduit Shannon, sur le conseil de Von Neumann, à nommer « entropie » sa définition de la quantité d'information[18]. On parlera donc indifféremment chez Shannon de la quantité d'information produite par une source de message ou de l'entropie de cette source de message. Le H du théorème de Shannon conjoint ainsi par un heureux hasard le H de Hartley (l'unité d'information, aujourd'hui oubliée) et le H de Boltzmann (l'interprétation statistique du concept thermodynamique d'entropie).

« Des quantités de la forme $H = -\Sigma p_i \log p_i$ jouent un rôle central dans la théorie de l'information […]. On reconnaîtra dans la forme de H celle de l'entropie définie dans certaines formulations de la mécanique statistique, où p_i est la probabilité d'être dans la cellule i de son espace de phase. H est alors, par exemple, le H du fameux théorème de Boltzmann. Nous appellerons $H = -\Sigma p_i \log p_i$ l'entropie de l'ensemble des probabilités p_i, …, p_n.[19] »

.

Si Shannon mentionne l'analogie entre formules sans s'avancer outre mesure sur la qualité réelle de l'analogie, rares seront ceux qui feront preuve de la même prudence. Ainsi, Warren Weaver cède à l'enthousiasme en sa préface, déclarant tout de go que « celui qui rencontre le concept d'entropie dans la théorie de la communication a de bonnes raisons de s'exciter – les bonnes raisons de celui qui a réussi à saisir quelque chose de fondamental et d'important.[20] » Parmi les cybernéticiens, Wiener ou Von Neumann, qui ne sont pas les moindres des mathématiciens du siècle passé, se rangeront parmi les soutiens déclarés d'une analogie qui est aujourd'hui largement périmée grâce aux avancées en informatique théorique et physique du calcul.

En admettant qu'il existe une analogie entre les formules – après tout la formule de la quantité d'information est suffisamment générale pour qu'elle se retrouve ailleurs qu'en physique statistique –, reste à justifier le passage de l'analogie formelle à une analogie réelle. Par quel miracle un nombre de digits binaires (quantité d'information) se laisserait-il convertir en un rapport entre quantité de chaleur et température (entropie) ? Seule l'intervention inopinée d'un démon pouvait prétendre dénouer l'affaire. Sur la suggestion de Wiener, on se

---

[18] Pour l'anecdote, [Denbigh, 1981] rapporte que Von Neumann aurait suggéré à Shannon l'emploi du terme entropie, non seulement en raison de la similitude entre les fonctions, mais aussi parce que « personne ne comprenant vraiment ce qu'est l'entropie, cela donnerait [à Shannon] avantage dans les débats. »
[19] [Shannon, 1949], pp. 50-51.
[20] [Shannon, 1949], p. 13.

mit donc à considérer que le vieux démon de Maxwell détenait la clé de la signification physique de l'information[21]. Reprenons quelques fils de cette histoire.

Dans ses *Leçons sur la théorie des gaz*, Boltzmann donne l'interprétation canonique du second principe de la thermodynamique, le principe de la croissance de l'entropie. L'entropie ($dS = dQ/dT$, où dQ est la variation de la quantité de chaleur et T la température) exprime en théorie de la chaleur l'irréversibilité des processus thermiques. Dans son mémoire de 1824, Sadi Carnot a montré qu'il existe un cycle thermique optimum, qui a comme propriété d'être réversible, à condition d'éviter les transferts de chaleur entre corps à températures différentes. C'est à cette idée que Rudolf Clausius donne forme mathématique dans les années 1850. Dans un cycle idéal de Carnot, où les échanges de chaleur se font donc à température constante, la variation d'entropie est nulle (transformation réversible), dans tous les autres cas, la variation d'entropie est toujours positive (transformation irréversible).

Boltzmann, en se fondant sur l'hypothèse atomique, fournit une explication statistique, au niveau microscopique, du phénomène macroscopique de croissance de l'entropie. On peut ainsi imaginer qu'à chaque état macroscopique, défini par des grandeurs mesurables comme la température ou la pression, correspond un ensemble de « complexions » microscopiques, définies par l'état singulier (position et vitesse) des molécules individuelles, que nous ne pouvons connaître dans le détail. Il est alors possible d'exprimer la loi de croissance de l'entropie comme « un théorème de probabilités[22] ». Si S est l'entropie, W la probabilité de l'état macroscopique ou le nombre de complexions susceptibles de le réaliser, alors $S = K \ln W$, où K est la constante de Boltzmann en joules par kelvin. L'entropie est ainsi le logarithme de la probabilité multiplié par la constante de Boltzmann. La formule de Boltzmann permet de comprendre la croissance de l'entropie comme la croissance de la probabilité de l'état macroscopique du gaz. Spontanément, par définition, un gaz dans un état macroscopique donné évoluera vers des états macroscopiques de plus en plus probables. A l'inverse, il est improbable que le système évolue vers des états moins probables. Croissance de la probabilité de l'état macroscopique et croissance de l'entropie expriment chacune à leur manière l'irréversibilité statistique du processus d'évolution du gaz. Mieux l'état le plus probable d'un système peut s'interpréter comme celui qui correspond au plus grand nombre de complexions microscopiques, soit l'état le moins ordonné et le plus homogène. On peut donc associer la croissance de la probabilité de l'état macroscopique exprimée par l'entropie avec l'évolution du système vers des états de plus en plus homogènes et de moins en moins ordonnés. L'entropie s'interprète alors comme une mesure du désordre moléculaire[23].

Entropie physique et quantité d'information s'expriment donc chacune comme un logarithme de probabilité ; reste à découvrir une situation physique dans laquelle les objets de la probabilité, dans un cas les micro-états, dans l'autre les messages, puisse faire l'objet d'une mesure commune. C'est précisément une telle situation que fournit opportunément l'expérience de pensée du démon de Maxwell. Maxwell, en cherchant à illustrer la valeur uniquement statistique de la seconde loi, - des phénomènes contraires à la seconde loi sont possibles, mais hautement improbables -, avait imaginé un dispositif qui aurait pour effet de violer le second principe de la thermodynamique, faisant passer sans travail de la chaleur d'un corps froid à un corps chaud.

[21] [Wiener, 1948], pp. 57-59.
[22] [Boltzmann, 1902], p. 58.
[23] Interprétation contestable et contestée. Voir, par exemple, [Tonnelat, 1982], qui conteste l'idée que l'état d'équilibre soit toujours l'état le plus probable. [Goldstein, 2001] donne une présentation claire de l'argument de Boltzmann et recense les différentes critiques qu'il a pu susciter.

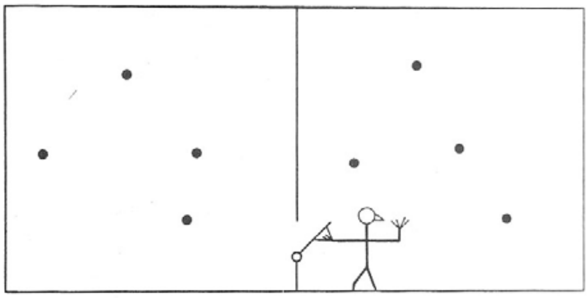

Fig. 2 – Le démon de Maxwell opérant le tri entre les molécules en fonction de leur vitesse. [Pierce, 1961], p. 199.

L'expérience de pensée est la suivante : un gaz remplit un récipient divisé en deux parties, de telle sorte que sa température soit plus élevée dans la partie A que dans la partie B, on peut alors concevoir qu'un petit être, capable d'agir sur les molécules individuelles, opère un tri parmi les molécules en fonction de leur vitesse, autorisant les molécules rapides de B (partie froide) à passer en A (partie chaude) et les molécules lentes de A à passer en B. La chaleur passerait alors spontanément et sans travail de la partie froide à la partie chaude. Autrement dit, si nous pouvions avoir accès aux propriétés des molécules individuelles, alors la seconde loi serait violée. Ce qui signifie dans l'esprit de Maxwell que la seconde loi ne vaut que dans la mesure où « le fait de s'occuper de larges masses de matière nous oblige à abandonner la méthode purement dynamique pour adopter une méthode statistique de calcul.[24] » Le démon intervient ainsi dans le débat sur la nature de la seconde loi et le conflit entre interprétation dynamique et statistique, mais il n'est jusqu'ici pas encore question d'information.

Pour comprendre la façon dont le démon est entré en cybernétique, il faut faire un dernier détour par les années 1920 et le travail de Leo Szilard. Dans son article de 1929, « De la diminution d'entropie dans un système thermodynamique par l'intervention d'êtres intelligents », Szilard cherche à montrer que le démon de Maxwell serait toujours empêché dans son action par le coût entropique des mesures qu'il doit effectuer afin de trier les molécules. Mieux, Szilard parvient dans le cas d'un gaz à une molécule à montrer que ce coût entropique des mesures doit nécessairement être supérieur ou égal à k log 2, si l'on veut sauver la seconde loi.

[24] James Clerk Maxwell, *Theory of Heat*, 1871, cité par [Leff and Rex, 1990], p. 4.

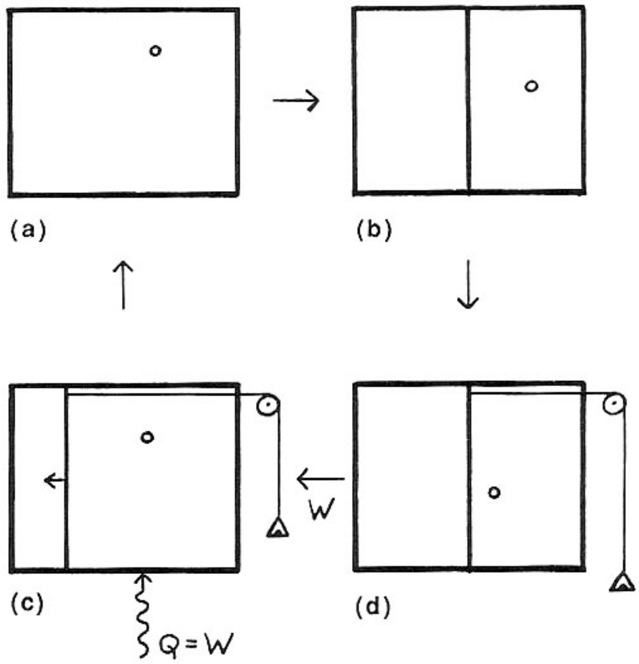

Fig. 3 – Les quatre temps du moteur de Szilard schématisés par [Leff and Rex, 1990], p. 15.

Szilard raisonne en imaginant un processus en quatre temps : (a) une molécule se trouve dans un cylindre, (b) cette molécule est piégée dans une moitié du cylindre par l'insertion d'une cloison qui fait office de piston, (d) le démon détermine dans quelle moitié du cylindre la molécule a été piégée et couple le piston à un poids qui produit un travail grâce aux mouvements de la molécule, (c) la pression du gaz fait se mouvoir le piston jusqu'au bout du cylindre où le gaz retrouve son état initial. Le processus peut recommencer (a).

Du point de vue de l'assimilation entre information et entropie, le « moteur de Szilard » produit deux résultats saisissants. Premièrement la quantité k log 2 à laquelle aboutit Szilard correspond, et ce dès les travaux d'époque de Nyquist et Hartley, à la quantité d'information impliquée dans un processus de choix binaire. Deuxièmement, le moteur suggère, lors du dernier temps, l'utilisation de la mesure, une diminution de l'entropie corrélative à celle de la quantité d'information. Notre connaissance quant à la position de la molécule au fur et à mesure de l'avancée du piston diminue, et tout se passe comme si notre information initiale avait été consommée pour produire le travail du piston et la diminution de l'entropie de l'environnement qu'elle implique. Le modèle du moteur de Szilard offre ainsi une situation privilégiée dans laquelle opérer une corrélation entre augmentation de l'information et diminution de l'entropie.

Information et entropie constituent ainsi deux concepts de probabilité que l'expérience de pensée du démon de Maxwell autorise à conjoindre en une seule et même situation physique. Tout se passe alors comme si nous pouvions assister à des jeux de conversion entre information et entropie. De là, deux question qui ne se recouvrent pas. Reste d'abord à savoir dans quelle mesure une telle construction des rapports entre information et entropie est légitime (il apparaît avec les progrès de la physique du calcul qu'elle conduit à une représentation partiale, sinon erronée, du problème[25]) ; mais reste aussi à savoir ce que

[25] En résumé, les travaux de Landauer ([Landauer, 1961]) et Bennet ([Bennett, 1973]) montrent que des opérations de copie réversible, assimilables à des opérations de prise d'information, sont concevables. Le principe de l'exorcisme du démon ne peut donc plus reposer sur le coût de la mesure, qui est dans certaines

peuvent bien signifier ces conversions d'entropie en information et vice versa. On aurait sans doute tort de croire qu'une réponse négative à la première question (l'analogie est-elle légitime ?) vide automatiquement la seconde (quelle interprétation pour l'analogie ?) de tout intérêt. En effet, les différentes versions de l'analogie dans les années 1950 n'apparaissent pas immunisées à part égale face aux critiques issues de la physique du calcul des années 1970. Si ces critiques sont majoritairement dirigées contre la version la plus développée de l'analogie, celle que l'on trouve chez Brillouin, elles laissent souvent de côté sa version cybernétique dont les divergences avec celle de Brillouin sont rarement perçues. Distinguer l'analogie à la manière de Wiener et à la manière de Brillouin permet de jeter un nouveau regard sur cette physique statistique enrichie aux dimensions de l'information que la cybernétique a prétendu incarner.

3. DEUX VERSIONS DE L'ANALOGIE INFORMATION-ENTROPIE OU DEUX INTERPRETATIONS DES PROBABILITES

Qu'est-ce qui distingue l'analogie information-entropie à la manière de Wiener et à la manière de Brillouin ? Sans doute une certaine représentation de ce que les probabilités signifient dans la définition de base des concepts d'information et d'entropie.

Publié en 1956, *La science et la théorie de l'information* rassemble les principales contributions de Brillouin sur la question, essentiellement des articles publiés depuis 1951 dans le *Journal of Applied Physics*. Le travail sera complété en 1964 avec *Scientific Uncertainty and Information* qui propose d'autres applications. Par rapport à ce que l'on trouve chez Szilard, le travail de Brillouin apporte deux nouveautés de taille. En 1929 Szilard ne disposait pas à proprement parler d'une définition de la quantité d'information. Tout ce que Szilard pouvait prouver est que les mesures opérées par le démon devaient avoir un coût minimal identifiable à k log 2. Dans l'article de Szilard, l'information n'est encore que le fantôme de l'entropie : si l'on veut sauver la seconde loi, alors on doit faire l'hypothèse que chaque mesure implique un coût entropique incompressible. Brillouin reprend la question, mais dispose, lui, d'un véritable concept d'information. Le livre et les articles de Brillouin proposent ainsi la mise en relation entre deux concepts pleins et non entre un concept (l'entropie) et son fantôme (la quantité qui sauve de manière ad hoc la seconde loi)[26]. Ce qui n'apparaît que comme hypothèse ad hoc chez Szilard devient chez Brillouin une authentique thèse : les deux quantités d'information et d'entropie désignent en réalité la même chose, au signe près. L'information devient dans la terminologie de Brillouin « néguentropie ».

Or, Brillouin, propose de fonder cette thèse, - deuxième apport -, sur les progrès de la physique quantique et de la théorie du rayonnement. Il fournit ainsi une explication en règle de la quantité minimale d'information. Pour mesurer, il faut produire ou utiliser un rayonnement, ce qui implique toujours un coût physique en termes d'entropie. Brillouin peut donc prétendre mettre en rapport information et entropie comme deux quantités distinctes au moyen d'une nouvelle interprétation du moteur de Szilard qui rend compte – ce que ne faisait pas Szilard – du coût incompressible de la mesure par la théorie du rayonnement.

---

situations privilégiées, nul. Ces travaux suggèrent que le coût irréversible apparaît lors des opérations de réinitialisation de la mémoire. La solution de Szilard ou Brillouin aurait donc le défaut de prendre la partie pour le tout : si certaines opérations de prise d'information sont bien coûteuses, toutes ne le sont pas ; la solution est donc ailleurs.

[26] C'est ce caractère ad hoc qui sert justement de cible à une première critique de l'analogie information-entropie chez Popper [Popper, 1957].

Mais à quoi sert le développement de l'analogie entre information et entropie ? L'analogie débouche chez Brillouin sur une théorie générale de la mesure physique, elle-même conçue comme argument décisif en faveur d'une philosophie opérationnaliste et idéaliste. Chez Brillouin, l'exorcisme du démon de Maxwell n'assure pas seulement la conversion de l'information et de l'entropie, mais de l'esprit et de la nature. Le raisonnement qui sous-tend tout le travail de *La science et la théorie de l'information* peut être reconstitué de la manière suivante. Un terme n'a de signification physique que si l'on est capable d'en produire une définition opérationnelle, c'est-à-dire d'exhiber un dispositif de mesure adapté. Or la théorie de l'information nous apprend que l'on ne peut jamais obtenir une mesure absolue. Toute mesure est toujours obtenue à un degré de précision finie. Une mesure infinie impliquerait un accroissement infini de l'entropie. Les termes physiques sont donc toujours relatifs à une situation d'observation, à un observateur muni de ses appareils aux conditions de précision limitées. Mais la théorie de l'information ne nous apporte pas seulement un résultat négatif (toute mesure possède un coût incompressible), elle permet d'intégrer ce résultat négatif à la théorie physique elle-même. Autrement dit, et c'est tout le sens de l'entreprise de Brillouin, on devient avec la théorie de l'information reformulée en théorie générale de la mesure physique capable de redéfinir l'ensemble des termes théoriques comme exprimant un état de connaissance incomplète. Brillouin nous enjoint ainsi à récrire l'ensemble de la physique en termes d'information, non plus comme une description objective et absolue des propriétés du monde tel qu'il va, mais comme une certaine expression de notre état nécessairement incomplet de connaissance.

« La théorie de l'information nous apprend qu'une expérience rigoureuse est absolument irréalisable, car elle donnerait une quantité d'information infinie, et exigerait une dépense infinie en néguentropie. Le fait nouveau c'est la considération du prix, du coût d'une observation. […] Nous utilisons les lois mathématiques pour ce qu'elles valent : ce sont des formules commodes pour résumer un matériel expérimental considérable, mais la valeur de ces formules ne doit pas être surestimée. Elles s'appliquent dans un domaine limité et avec une précision limitée.[27] »

Avec Brillouin ce n'est donc pas seulement la théorie de l'information qui rejoint la physique, mais bien plutôt toute la physique qui en vient à s'exprimer en termes d'information. L'information devient un terme physique au moment où toute la physique se révèle affaire d'information.

Un rapide survol des textes de Wiener sur la question suffit à donner une idée du malentendu. Si nous regardons le passage consacré au démon placé à la conclusion du deuxième chapitre de *Cybernetics*, nous voyons qu'il s'agit chez Wiener de toute autre chose qu'une théorie de la mesure physique. La question posée par Wiener porte, à la suite du *What is Life?* de Schrödinger, sur les rapports entre théorie physique et science du vivant. Qu'est-ce que les lois physiques, en particulier les lois de la physique statistique, peuvent nous apprendre sur les êtres vivants ? Wiener est confronté à une objection : l'étude des systèmes thermodynamiques porte habituellement sur des systèmes à l'équilibre ou proche de l'équilibre, ce qui n'est manifestement pas le cas des vivants. Tout l'enjeu de la discussion consiste à promouvoir un nouveau type d'explication physicaliste des phénomènes biologiques, distinct des explications énergétistes traditionnelles. On reconnaît ici un des points clés de la première cybernétique, qui s'attache à distinguer les transferts d'énergie et les transferts d'information, afin de promouvoir ces derniers comme modèles explicatifs. La situation du démon de Maxwell offre alors l'occasion de nouer pour la cybernétique physique statistique, théorie de l'information et biologie.

---

[27] [Brillouin, 1956], p. 285.

La solution de Wiener diffère manifestement de celle de Brillouin. Le modèle n'est d'abord pas le même puisque Wiener utilise une description de la situation du démon semblable à celle donnée par Maxwell dans la *Théorie de la chaleur*. Nous avons affaire à un démon qui trie les molécules en fonction de leur vitesse. Contrairement à ce que l'on trouve chez Brillouin ou Szilard, il n'est donc pas question chez Wiener d'un démon de pression, d'un gaz à une seule molécule, ou même de l'intelligence du démon.

Le principe de la solution proposée par Wiener repose sur l'idée d'un couplage inévitable entre le démon et le gaz. Comme chez Brillouin, ce couplage résulte de la prise d'information. Le transfert d'information a beau avoir lieu à des niveaux d'énergie très faibles, il n'en reste pas moins que au final, sur le long terme, le démon finira par être lui aussi gagné par le bruit, soumis aux fluctuations aléatoires du gaz. Si le démon peut bien fonctionner pour un temps, il finit donc peu à peu par sombrer avec son environnement. Le démon cybernétique se mue alors en créature leibnizienne, étourdie par une multitude de perceptions évanouissantes.

« Sur le long terme, le démon de Maxwell est sujet à un mouvement aléatoire fonction de la température de son environnement et comme Leibniz le dit de ses monades, il reçoit énormément de petites impressions qui le plongent dans « une sorte de vertige », le rendant incapable de perceptions claires. Il cesse alors d'agir comme un démon de Maxwell.[28] »

Si la solution de Wiener comporte comme celle de Brillouin la question du coût de la mesure, elle lui assigne cependant une place tout à fait différente. Le coût de la mesure, pivot de la démonstration chez Brillouin, ne constitue chez Wiener qu'un élément annexe, qui n'est pas élaboré pour lui-même. C'est que le problème de Wiener n'est pas tant d'exorciser le démon au moyen d'une philosophie opérationnelle de la physique, que de montrer que les démons existent, qu'ils sont parmi nous, réduisant pour un temps et pour un lieu l'accroissement inéluctable de l'entropie. Le démon de Wiener n'est pas l'observateur modeste de Brillouin, mais l'enzyme, la cellule ou l'organisme. Le démon est à chasser du côté de ces systèmes biologiques métastables capables de renverser pour un temps le cours de l'entropie et de maintenir leur ordre à travers et par le désordre.

« Il n'y a aucune raison de supposer que des démons métastables ne puissent exister en réalité ; de fait, les enzymes sont peut-être bien des démons de Maxwell métastables, diminuant l'entropie, sans doute pas en séparant les particules lentes des particules rapides, mais par quelque autre processus équivalent. Nous pouvons aussi bien regarder les organismes vivants, comme l'Homme lui-même, sous ce jour.[29] »

A contrario, l'interprétation biologique du démon et de l'analogie information-entropie ne reçoit aucun traitement dans le livre de Brillouin[30]. La relation entre entropie, information et organisation est ramenée à quelques pages en fin d'ouvrage sous la rubrique « problèmes divers ».

Il nous paraît évident que l'analogie entre information et entropie fonctionne à front renversé chez Wiener et chez Brillouin. Wiener cherche à penser l'information comme une quantité physique, là où Brillouin cherche à penser les quantités physiques comme de l'information. L'analogie entre information et entropie peut donc nous conduire à deux représentations tout à fait opposées de l'information, selon qu'on se représente l'information

[28] [Wiener, 1948], p. 58.
[29] Ibid.
[30] Il y a ici une rupture entre le premier travail de Brillouin ([Brillouin, 1949]), qui est directement inspiré du *Cybernetics* de Wiener et du *What is Life ?* de Schrödinger, et les travaux suivants, à partir de 1950, qui délaissent la question du vivant et de son organisation pour la question du coût de la mesure en physique. Nous passons alors d'une question « objective », au sens où elle porte sur une propriété des vivants (l'organisation) que l'on suppose indépendante de l'observateur et que l'on cherche à quantifier – ce qui sera le programme de la théorie des automates chez Von Neumann –, à une question « subjective », au sens où elle porte sur l'opération de prise de connaissance, envisagée du côté du sujet.

comme l'expression d'une propriété objective, le degré d'organisation (Wiener) ou l'expression d'une propriété subjective, le degré de précision de notre connaissance expérimentale (Brillouin). Information s'entend tantôt au sens de la prise de forme, de l'informé, tantôt au sens de la prise de connaissance, de l'informatif. Cette dualité ne tombe pas du ciel sur un concept d'information qui n'en demandait pas tant, elle poursuit et amplifie au contraire une ligne de clivage au sein du concept mathématique d'information. En choisissant de nommer information la fonction H, puis en y accolant le terme d'entropie, Shannon ou Wiener, se sont installés délibérément sur la ligne de faille entre probabilité épistémique et probabilité fréquentiste. Qu'est-ce qui relève des probabilités épistémiques dans la définition de Shannon ? Rien, ou presque rien. Aucun usage de la règle de Bayes, aucune quantification des degrés de croyance. Et pourtant de quoi peut-il s'agir avec de l'information si ce n'est des degrés de connaissance. L'incertain pour le destinataire n'est-il pas le plus informatif ? Mais si l'on en revient maintenant à la construction mathématique, nous n'avons plus à faire qu'à une étude des propriétés statistiques de générateurs aléatoires, rapportées au problème du codage et de la construction des filtres. Le concept d'information met bout à bout, sans aucunement résoudre ni même expliciter la tension, mesure de la dispersion statistique d'une source ergodique et mesure du degré de connaissance, mesure de la forme ou de la structure et mesure de l'informatif. Cette confusion entre l'épistémique et le stochastique réside au principe même du concept d'information, dans le postulat qui opère la conversion entre l'intuitif et le formel, affirmant que plus c'est incertain, au sens de la dispersion statistique, plus c'est informatif, au sens ordinaire du mot qui implique la connaissance. Le concept d'information s'installe précisément sur cette dualité entre incertitude subjective et objective, avec la promesse non explicitée d'une conversion naturelle et immédiate entre structure statistique et degré de croyance ; rien moins qu'une théorie (mal) unifiée de la probabilité. Si une telle description est exacte, on comprend aisément que le concept d'information ait pu donner prise à des interprétations aussi radicalement opposées que celles de Wiener et Brillouin, retombant tantôt du côté objectif de la probabilité, tantôt du côté subjectif.

Reste enfin à montrer, afin de fortifier l'hypothèse, comment la représentation des probabilités peut intervenir dans les interprétations divergentes de l'information chez Wiener et Brillouin. Le cas de Brillouin est exemplaire. En un sens l'analogie entre information et entropie ne constitue qu'un cas particulier par rapport à la situation de l'ensemble des concepts physiques qui doivent pouvoir s'exprimer en termes de mesure et d'information. Mais d'un autre côté, c'est bien l'analogie entre information et entropie qui sert de pivot à la l'extension de la théorie de l'information à l'ensemble de la physique. Or, si nous observons en détail la démonstration de Brillouin, celle-ci repose à l'évidence sur une lecture qui est déjà subjective de l'entropie. Autrement dit, la philosophie de la physique qui apparaît fondée sur l'exorcisme du démon est en réalité déjà requise par l'exorcisme. La démonstration présuppose ce qui est à démontrer.

L'association de l'information et de l'entropie repose en effet chez Brillouin sur une interprétation déjà « informationnelle » de l'entropie entendue comme l'expression d'un défaut de connaissance de notre part quant à la structure du système physique observé. Selon cette interprétation le fait que l'entropie d'un système soit élevée exprime le fait que la connaissance que nous pouvons avoir de son état réel, microscopique, présente d'importantes lacunes, puisque le nombre d'états microscopiques possibles compatibles avec l'état macroscopique observé est élevé. A l'inverse une information élevée quant à l'état du système correspond à une entropie faible (le nombre de complexions possibles est moindre, nous en savons plus).

« L'entropie est en général considérée comme exprimant l'état de désordre d'un système physique. D'une façon plus précise, on peut dire que *l'entropie mesure le manque d'information* sur la véritable structure du système. Ce manque d'information implique la possibilité d'une grande variété de structures microscopiques distinctes qui sont, en pratique, impossibles à distinguer les unes des autres.[31] »

Cette définition « subjective », qui renvoie l'entropie à un degré de connaissance, est une manière d'interpréter (fort répandue, au demeurant) l'apparition des probabilités dans la définition de Boltzmann[32]. Le raisonnement est le suivant : s'il y a des probabilités, c'est qu'il y a un état de connaissance incomplète, l'entropie, qui est un concept de probabilité, doit donc s'interpréter comme l'expression d'un degré de connaissance[33]. De l'interprétation épistémique spontanée des probabilités engagées dans la définition de l'entropie se déduit immédiatement chez Brillouin la conversion entre entropie et information. Soit, dans une formule particulièrement malheureuse, « l'entropie d'un système décroît lorsqu'on obtient de l'information, car cette information réduit le nombre des complexions.[34] » Tout cela n'est évident qu'à la double condition de présupposer non seulement que l'entropie exprime naturellement un état de connaissance, mais aussi que l'information s'entend au sens ordinaire de ce qui est plus ou moins informatif.

Renversons les présupposés, interprétons de manière objective information, entropie et probabilité, et nous obtenons la version de Wiener. L'entropie devient une mesure du désordre, qui n'est pas modifiée par la connaissance que nous pouvons en avoir, l'information mesure symétrique du degré d'ordre. D'où procède un tel renversement ? Manifestement de la représentation des probabilités engagées dans la définition de l'entropie et de l'information. L'usage des probabilités en physique statistique doit-il s'interpréter pour Wiener comme l'expression d'une connaissance incomplète ? Nullement ; les probabilités sont toujours reliées chez Wiener au hasard, conçu comme terme physique, à l'existence avérée de séries temporelles aléatoires. Les formules imagées de *The Human Use of Human Beings* sont on ne peut plus claires sur ce point.

« Le hasard a été admis, non pas simplement comme un instrument mathématique pour la physique, mais comme une partie de la chaîne et de la trame [de l'univers]. [Il faut reconnaître] un élément fondamental de hasard dans la texture de l'univers lui-même.[35] »

Si Brillouin rapporte les probabilités à la situation du sujet connaissant, Wiener n'y voit qu'une conséquence de l'existence de séries temporelles aléatoires. Le reste – la représentation subjective ou objective de la quantité d'information – s'ensuit.

---

31 [Brillouin, 1956], p. 155.

32 Pour une critique de ce point de vue, [Popper, 1974], [Denbigh, 1981] ou encore [Goldstein, 2001].

33 Est-ce à dire qu'une interprétation subjective des probabilités mène nécessairement à une philosophie subjectiviste de la physique ? Certainement, non. Ce point est défendu de manière convaincante par [Bricmont, 2001], p. 6. Il n'en reste pas moins que la connexion existe, à un état particulièrement pur, chez Brillouin… ce que déplore évidemment Bricmont – « un tenant de l'interprétation subjective des probabilités peut très bien soutenir qu'il existe des faits et des lois objectifs dans le monde et considérer les probabilités comme un simple outil adapté aux situations où notre connaissance de ces faits et lois est incomplète. […] Mais la position du subjectivisme philosophique commence souvent par confondre le monde et notre connaissance (par exemple, le discours vague sur le fait que tout est information oublie souvent le fait que « l'information » est toujours information à propos de quelque chose qui n'est pas information). Ainsi, on ne doit pas faire des Bayesiens les compagnons de route naturels des subjectivistes en philosophie. »

34 [Brillouin, 1956], p. 148. La formule est malheureuse en ce qu'elle semble affirmer que le savoir (l'information) agit directement sur l'état du système physique, « réduisant le nombre de complexions ». De là, un torrent de critiques contre ce subjectivisme qui croit mouvoir les machines avec du savoir mais ne produit que de l'air chaud et de l'entropie en proportion de son non-savoir (la formule se retrouve dans les deux textes de Popper, [Popper, 1957] et [Popper, 1974]). Le problème de Brillouin est plutôt d'intégrer les conditions de précision expérimentale dans l'expression des termes théoriques.

35 [Wiener, 1954], p. 11.

CONCLUSION

Il apparaît donc au terme de ce développement que l'usage des probabilités dans la définition mathématique de l'information donne lieu à deux lectures opposées de ce que l'information peut signifier en réalité, dès lors que l'on s'éloigne du strict domaine de l'ingénierie des télécommunications. La quantité d'information peut s'interpréter comme mesure du degré d'ordre – lecture qui sera retenue dans les théories algorithmiques qui approfondiront le rapport entre information et complexité – ou mesure du degré de connaissance. Cette dualité entre l'informé et l'informatif se nourrit de l'opposition entre interprétation épistémique et fréquentiste des probabilités mobilisées par la définition de la quantité d'information. Cette dualité apparaît au grand jour dans la question du statut physique de l'information, rapportée à l'expérience de pensée du démon de Maxwell.

Cette histoire suggère trois remarques en guise de conclusion. Premièrement, les oppositions tranchées quant à l'interprétation physicaliste de l'information offrent une sorte d'application dérivée à la distinction entre probabilité épistémique et fréquentiste. Il semble que l'on ait là un terrain sur lequel cette dualité produit des effets, en dehors du développement de la théorie des probabilités proprement dite. Deuxièmement, rapporter les divergences d'interprétation de la quantité d'information à un clivage dans l'interprétation des probabilités permet de donner de la consistance à cette zone intermédiaire où se négocie la signification de la formule de la quantité d'information. Ces discussions n'ont rien de trivial ; non seulement au sens où les problèmes envisagés quant au statut des probabilités ne le sont pas, mais aussi au sens où elles produisent des effets, ici en aval de la définition de la quantité d'information, pour ce qui est de la position du programme cybernétique, la définition des questions urgentes ou des pistes à explorer. La question de la signification de l'information participe à plein au montage du programme interdisciplinaire de la cybernétique dans ses plus riches années. Sous prétexte qu'elle ne relève pas d'une rationalité effective et calculatoire, mais d'une raison spéculative, on aurait sans doute tort de la rejeter hors de la rationalité tout court, pour la projeter du côté des fausses sciences et des idéologies scientifiques. Enfin, troisièmement, revenir à cette préhistoire méconnue des rapports entre information et physique permet d'éclairer certaines des prises de position les plus radicales de la physique du calcul des années 1970, qui doit sans doute beaucoup plus à la cybernétique qu'il n'y paraît, ainsi que de jeter un pont entre la théorie de l'information à la Shannon-Wiener et la théorie algorithmique qui reprend le flambeau de l'organisation et de la complexité, poursuit l'intuition de Wiener d'une convergence fondamentale entre étude de la probabilité et étude de l'information. Il y a sans doute là des chemins continus d'investissement théorique qui n'ont pas encore tout à fait été explorés.

BIBLIOGRAPHIE

BENNETT C., « Logical reversibility of computation », *IBM J. Res. Dev.*, 17, 1973, pp. 525-532.

BENNETT C., « The thermodynamics of computation – a review », *Int. J. Theor. Phys.*, 21, 1982, pp. 905-940.

BOLTZMANN L., *Leçons sur la théorie des gaz*, (trad. fr. A. Galloti, Paris, Gabay, 1987 [1902]).

BRICMONT J., « Bayes, Boltzmann and Bohm: Probabilities in Physics », in *Chance in Physics, Foundations and Perspectives*, Berlin, Springer, 2001, pp. 3-24.

BRILLOUIN L., *La science et la théorie de l'information*, Paris, Gabay, 1988 [1956].

BRILLOUIN L., *Scientific Uncertainty and Information*, New York, Academic Press, 1964.

CARNOT S., *Réflexions sur la puissance motrice du feu et sur les machines propres à développer cette puissance*, Paris, Gabay, 1990 [1824].

CHAITIN G., *Algorithmic Information Theory*, Cambridge, Cambridge University Press, 2004 [1987].

DELAHAYE J-P., *Information, complexité, hasard*, Paris, Hermès, 1999.

DENBIGH K., « How subjective is entropy ? », *Chem. Brit.*, 17, 1981, pp. 168-185.

DESROSIÈRES A., « Les recherches de Ian Hacking sur l'histoire des usages des probabilités et des statistiques dans le raisonnement inductif », *Journ@l électronique d'Histoire des Probabilités et de la Statistique/ Electronic Journal for History of Probability and Statistics*, Vol.2, n°1. Juin/June, 2006.

GOLDSTEIN S., « Boltzmann's Approach to Statistical Mechanics », in *Chance in Physics, Foundations and Perspectives*, Berlin, Springer, 2001, pp. 39-54.

HACKING I., *The Emergence of Probability* (trad. fr. M. Dufour, *L'émergence de la probabilité*, Paris, Seuil, 2002 [1975]).

HACKING I., *The Social Construction of What?* (trad. fr. B. Jurdant, *Entre science et réalité, la construction sociale de quoi ?*, Paris, La Découverte, 2001 [1999]).

HACKING I., *An Introduction to Probability and Inductive Logic* (trad.fr, M. Dufour, *L'ouverture au Probable, Eléments de logique inductive*, Armand Colin, Paris, 2004 [2001]).

HARTLEY R., « Transmission of Information », *Bell System Technical Journal*, 7, 1928, pp. 535-563.

HEIMS S., *John Von Neumann and Norbert Wiener, From Mathematics to the Technologies of Life and Death*, Cambridge, MIT Press, 1980.

LANDAEUR R., « Irreversibility and heat generation in the computing process », *IBM J. Res. Dev.*, 5, 1961, pp. 183-191.

LEFF H. AND REX A. (ed.), *Maxwell's Demon, Entropy, Information, Computing*, Princeton, Princeton University Press, 1990.

MARTIN T., « Logique du probable de Jacques Bernoulli à J.-H.Lambert », *Journ@l électronique d'Histoire des Probabilités et de la Statistique/ Electronic Journal for History of Probability and Statistics*, Vol.2, n°1b, Novembre/November 2006.

MASANI P., *Norbert Wiener 1894-1964*, Bâle, Birkhäuser Verlag, 1990.

MINDELL D., *Between Human and Machine, Feedback, Control, and Computing before Cybernetics*, Baltimore, John Hopkins University Press, 2002.

NYQUIST H., « Certain Factors affecting Telegraph Speed », *Bell System Technical Journal*, 3, 1924, pp. 324-346.

PATY M., *Einstein philosophe*, Paris, Puf, 1993.

PIERCE J., *Symbols, Signals and Noise : The Nature and Process of Communication*, New York, Harper & Brothers, 1961.

POPPER K., « Irreversibility; Or, Entropy since 1905 », *The British Journal for the Philosophy of Science*, 8, 30, Aug. 1957, pp. 151-155.

POPPER K., Unended quest, Londres, Routledge, 2002 [1974].

ROSENBLUETH A., WIENER N. AND BIGELOW J., « Purpose, Behavior and Teleology », *Philosophy of Science*, 10, 1943, pp. 18-24.

SCHRÖDINGER E., *What is Life ?*, Cambridge, Cambridge University Press, 1944.

SEGAL J., *Le Zéro et le Un, Histoire de la notion scientifique d'information au 20e siècle*, Paris, Syllepse, 2003.

SHANNON C., *The Mathematical Theory of Communication*, Chicago, University of Illinois Press, 1949.

SHANNON C., « The Bangwagon », *IRE Transactions on Information Theory*, 2, mars 1956, p. 3.

SZILARD L., « On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings », *Behavioral Science*, 9, 1964 [1929], pp. 301-310.

TONNELAT J., « The organization and evolution of living systems as normal consequences of the laws of thermodynamics », *Advances in Applied Probabilities*, 14, 2, 1982.

VON PLATO J., *Creating Modern Probability, Its Mathematics, Physics and Philosophy in Historical Perspective*, Cambridge, Cambridge University Press, 1994.

WIENER N., *Cybernetics, or control and communication in the animal and the machine*, Cambridge, MIT Press, 2000 [1948].

WIENER N., *Times Series,* Cambridge, MIT Press, 1949.

WIENER N., *The Human Use of Human Beings, Cybernetics and Society*, London, Free Association Books, 1989 [1954].

WIENER N., *I am a Mathematician, The Latter Life of a Prodigy*, New York, Doubleday, 1956.