



Le concept d'information chez Shannon et Wiener

Mathieu Triclot

► To cite this version:

Mathieu Triclot. Le concept d'information chez Shannon et Wiener. Séminaire science, légitimité, médiation équipe Maaticah, Jun 2008, Paris, France. halshs-00514076

HAL Id: halshs-00514076

<https://shs.hal.science/halshs-00514076>

Submitted on 2 Jan 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Le concept d'information chez Shannon et Wiener

Mathieu Triclot

Séminaire Maaticah, Paris 8

Cette intervention a pour objet « le concept d'information chez Shannon et Wiener », avec comme objectif d'essayer de faire le point sur les éventuelles différences d'approche entre Shannon et Wiener concernant l'information.

0.1

Le problème est simple. On voit apparaître, coup sur coup, à l'été et à l'automne 1948 deux textes qui proposent une mesure de la quantité d'information.

Nous avons d'abord les célèbres articles de Claude Shannon dans le Bell System Technical Journal, dans la livraison de juillet et octobre 1948, qui proposent « Une théorie mathématique de la communication ».

Puis, toujours en octobre 1948, nous avons la publication de ce best-seller inattendu de la littérature scientifique qu'est le *Cybernetics* de Norbert Wiener.

Dans ces deux textes, nous trouvons une mesure de la quantité d'information. Cette mesure est globalement reconnue comme équivalente, à commencer par les deux principaux protagonistes.

Ainsi, on lit sous la plume de Wiener dans *Cybernetics* que – je cite – « Pour prendre en compte cet aspect de l'ingénierie des communications, il nous a fallu développer une théorie statistique de la quantité d'information, dans laquelle l'unité d'information correspondait à ce qui est transmis par une décision entre deux alternatives également probables. Cette idée est apparue à peu près au même moment chez plusieurs auteurs, parmi eux le statisticien R.A. Fisher, le Dr. Shannon des Laboratoires Bell et l'auteur. » La mention de Fisher est bizarre, la chronologie fantaisiste ; toujours est-il que Shannon est cité, et ce n'est pas la seule mention. Ici, il s'agit d'un extrait de l'introduction du livre ; mais nous trouvons deux mentions de Shannon dans le chapitre consacré à la mesure de l'information, dont l'une qui mentionne un résultat commun – je cite – « qui a déjà été obtenu par l'auteur et Shannon pour le taux de transmission de l'information ».

Shannon, n'est pas en reste, puisque nous trouvons plusieurs mentions extrêmement favorables du travail de Wiener, qui vont jusqu'à une forme de reconnaissance de priorité.

Ainsi, je cite : « La théorie de la communication doit énormément à Wiener pour une part importante de sa philosophie de base et de sa théorie. Son rapport classique au NDRC, *The interpolation, Extrapolation, and Smoothing of Stationary Time Series* (Wiley, 1949), contient la première formulation explicite (*clear-cut*) de la théorie de la communication comme un problème statistique, l'étude des opérations sur des séries temporelles. Ce travail, bien que portant au premier chef sur la prédiction linéaire et le problème des filtres, constitue une référence parallèle importante en lien avec le présent article. Nous pouvons aussi nous référer au *Cybernetics* de Wiener (Wiley, 1948), qui porte sur les problèmes généraux de la communication et du contrôle. »

S'ajoute une autre mention dans les remerciements à la fin de l'ouvrage, dans laquelle on voit Shannon remercier ses collègues des Bell Labs, Bode, Pierce, McMillan et Oliver et ajouter que – je cite – « la solution élégante par le Professeur N. Wiener des problèmes du filtrage et de la prédiction d'ensembles stationnaires a considérablement influencé la réflexion de l'auteur dans ce champ. »

0.2

J'avais un doute sur le fait que ces références à Wiener apparaissent bien dans le texte des articles de 1948, et non seulement dans la version republiée sous forme de livre en 1949. La première mention comporte en effet des références à des textes non encore parus en 1948, *Cybernetics* bien sûr, mais aussi le rapport au NDRC qui ne sera republié qu'au mois d'août 1949. De plus la réédition des articles sous forme de livre en 1949 explique que les articles ont été republiés tels quels, à l'exception de quelques corrections mineures et de l'ajout de références additionnelles. Après vérification, le doute est levé.

Les références à Wiener sont déjà là dans l'article de 1948, avec juste des corrections mineures : « His classic NDRC ... to appear soon in book form. Wiener's forthcoming book "Cybernetics" dans la version de 1948. »

Si j'ai pu avoir des doutes, c'est que le consensus affiché en 1948 ou 1949 ne va pas sans quelques tensions sur les questions de priorité entre le professeur de mathématiques du MIT et le chercheur des Bell Labs ; et l'on trouve de part et d'autre quelques formules pleines de fiel. Nous y reviendrons.

Quoi qu'il en soit, pour l'instant, nous sommes fondés à reconnaître l'équivalence des mesures de Shannon et Wiener. On pourrait citer d'autres sources, par exemple, la recension favorable que Shannon donne de *Cybernetics* pour la revue des ingénieurs radio en 1949.

De fait, les définitions ont été reconnues équivalentes, au-delà des protagonistes eux-mêmes, par les intervenants extérieurs. Je pense, par exemple, à ce que peut dire Donald MacKay dans la 8^e conférence Macy de 1951 qui range sous le concept d'information de sélection les mesures de Shannon et Wiener, mais aussi aux travaux de Marcel-Paul Schützenburger en France qui ont été étudiés par Jérôme Segal et qui abordent directement la question de l'équivalence des définitions de Shannon, Wiener et Fisher, au sein d'une expression qui serait plus générale.

0.3

Une fois que l'on a dit cela, il semble du coup, que l'objet de cette communication s'évanouisse : à quoi bon se poser la question des différences dans la construction d'une mesure de l'information chez Wiener et Shannon, dans le contexte américain, dès que lors que les mesures sont, sur le fond, sinon sur la forme, équivalentes ?

Où passe la différence si sur le fond il y a identité ? Est-ce que l'on ne sera pas amené à ne pointer que des différences superficielles, comme celles qui ont trait à ces petites querelles de priorité dont j'ai fait mention tout à l'heure et que l'on voit ressurgir dans la littérature secondaire consacrée à la question ?

Je dois donc justifier un peu la question que je vous pose, qui est celle des différences d'approche entre Shannon et Wiener.

Je dois d'abord vous avouer que cette question des différences éventuelles entre l'information de Shannon et celle de Wiener est arrivée pour moi comme une question adjacente, un chapitre dans un problème plus global, qui était celui que je posais dans la thèse, qui était (a) de savoir ce que l'information voulait dire pour les protagonistes de la cybernétique, quelle représentation ils se faisaient de cette nouvelle « chose » entre guillemets qui est apparue comme information après la guerre, (b) et – question additionnelle – de se demander en particulier si ces représentations de ce que ça peut être que de l'information sont bien unifiées. Tout le monde se met à parler d'information, mais qu'est-ce que l'on veut dire par là, qu'est-ce qui est visé ?

Or, il me semble qu'on a de bonnes raisons de penser si l'on regarde ce qui est visé sous le nom d'information dans d'autres domaines, et je pense en particulier à tous ces textes de la même période sur le rapport entre les calculateurs et la pensée, que l'information n'apparaît pas sous une forme univoque. Qu'est-ce que c'est que de l'information ?

Il me semble que si l'on veut résumer cela au plus saillant, on peut se dire que sous le vocable d'information on peut viser soit une suite de symboles, les fameux digits binaires 0 ou 1 que l'informatique a fini par faire rentrer dans notre vie quotidienne, soit disons un signal, c'est-à-dire l'expression d'un ordre, d'un agencement matériel.

Mais, signal ou symbole, autrement dit analogique ou numérique, continu ou discret, on sait convertir l'un en l'autre. Techniquement, on peut se dire que la distinction n'a pas grande importance. Et cela fait partie des arguments que l'on retrouve dans les conférences Macy. N'érigeons pas la distinction entre numérique et analogique en absolu répètent Wiener ou Von Neumann.

De fait, si on sait échantillonner un signal, on peut passer du signal au symbole, du continu au discret et vice versa. Et d'ailleurs le premier dispositif technique mentionné par Shannon dans l'article de 1948, le système PCM, de modulation par impulsion codée consiste précisément à transmettre un signal continu sous forme numérique, ce qui a un intérêt pour les économies qu'on peut réaliser par un codage efficient ou bien à l'inverse pour l'introduction d'une clé de chiffrement en cryptographie.

Reste que je crois que ce partage entre le symbole et le signal n'est pas totalement dénué de sens, si on se réfère aux débats du côté de ce qui s'appellera un peu plus tard l'intelligence artificielle, sur les rapports entre les nouveaux calculateurs et la pensée.

Qu'est-ce que je veux dire par là ? Il me semble que ce qui caractérise la cybernétique, disons en tant que programme de recherche, en tant que pari pris sur l'avenir, sur ce qui est intéressant du point de vue de la recherche, c'est le fait que la cybernétique a pris parti pour une représentation physicaliste, matérialiste de l'information. Autrement dit, il me semble que chez les cybernéticiens, nous avons cette idée que l'information comme suite de symboles constitue une réalité dérivée et que ce qui compte de manière plus fondamentale, c'est disons le signal, c'est-à-dire l'information en tant que grandeur physique.

Je crois que cette priorité là s'observe très bien dans la manière dont Wiener introduit la cybernétique comme un département de la physique. Qu'est-ce que c'est que la cybernétique pour Wiener ? C'est une physique de l'information ; physique de l'information qui constitue elle-même un chapitre de la mécanique statistique.

Mais je crois que cette distinction entre le signe et le signal est aussi extrêmement opératoire si l'on veut comprendre disons les querelles d'héritage autour de l'œuvre de Warren McCulloch et Walter Pitts. La plupart des tenants de l'Intelligence Artificielle symbolique, au premier rang desquels Marvin Minsky, qui est lui-même un étudiant de McCulloch, vont choisir de faire de McCulloch, non pas un neurobiologiste, mais un logicien.

Autrement dit, ils vont lire un article aussi important que l'article de 1943, sur « un calcul logique des idées immanent à l'activité nerveuse », comme un article de logique formelle, une théorie formelle des automates binaires à seuil. Or, si on regarde le travail de McCulloch et Pitts, mais surtout la lecture critique qu'en donne quelqu'un comme Von Neumann, l'accent ne porte pas du tout sur un travail qui serait purement logique, manifestement ce qu'ils ont tête c'est d'essayer de décrypter la façon dont notre logique peut fonctionner au niveau du cerveau. Et là on trouve chez Von Neumann une proposition qui est constamment répétée : notre logique, symbolique, n'est – je cite – qu'un code court, métaphore informatique, exprimant de manière simplifiée un code long qui serait en quelque sorte le langage du cerveau, qui est lui massivement parallèle et stochastique.

Autrement dit, l'information comme symbole n'est qu'une réalité dérivée par rapport à quelque chose de plus fondamental qui est le traitement de l'information tel qu'il s'opère au niveau physique de base.

Autrement dit, il y a convergence – qui ne va pas sans tiraillements, ils ne sont pas d'accord sur tout, mais – convergence entre le projet d'une physique statistique de l'information à la manière de Wiener et ce que Von Neumann cherche à faire avec sa théorie statistique des automates.

Du coup, on est obligé de tenir à la fois que le numérique et l'analogique peuvent se convertir l'un en l'autre, ce qui ne veut pas dire qu'il y ait les mêmes propriétés au niveau technique, en termes de fiabilité ou de degré de précision pour les calculateurs, par exemple ; conversion réciproque, mais aussi du côté de la cybernétique, l'affirmation répétée d'un primat du signal physique sur le code symbolique, qui n'en est que l'expression simplifiée.

Je pense en particulier à l'échange entre Wiener et Von Neumann lors de la sixième conférence Macy en 1950 dans laquelle on voit Wiener exposer le projet « d'une physique des dispositifs numériques », consistant à ramener le comportement numérique des composants à des effets de seuil sur une dynamique sous-jacente d'ordre continu. L'idée est que les états digitaux ou discrets correspondent à des états d'équilibre qui demeurent la plupart du temps

insensibles aux petites variations continues, mais qui peuvent basculer vers d'autres états d'équilibre à partir du moment où les variations dépassent un certain seuil. Von Neumann reprend l'idée à son compte lors de la conférence suivante.

0.4

Vous avez donc toutes les cartes en main pour juger de la question que je pose aujourd'hui. Est-ce que cette différence d'approche de l'information se retrouve dans l'autre champ qui est décisif pour l'émergence de l'information comme objet de science, l'ingénierie des communications, avec cette fameuse mesure de la quantité d'information que l'on va trouver chez Shannon et Wiener ?

Une fois reconnue l'équivalence formelle des définitions, est-ce qu'on peut cependant se poser la question de savoir ce qui est visé par ces définitions, ce que c'est que de l'information chez l'un et l'autre ? Quels sont les objets dont l'information est une mesure ?

Il me paraît certain que l'on peut tout à fait se passer de cette question de ce qui est représenté par la mesure de l'information et se contenter d'appliquer la formule. Du moment que cela marche, comme me disent mes étudiants, à quoi bon se casser la tête avec de la philosophie et la question de ce que cela peut vouloir dire au-delà de l'application. Reste que cette question là, il me semble que la cybernétique se l'est posée et qu'elle n'est pas totalement dénuée de sens.

Du coup, cela oblige aussi à revenir aux contextes dans lesquels les définitions de Wiener et Shannon ont été produites. Qu'est-ce qui est nommé information dans un cas et dans l'autre ?

Je vous propose donc de parcourir les pièces de ce dossier déjà bien balisé, en commençant par examiner les propriétés des définitions produites, puis en revenant à l'histoire parallèle de la constitution de ces deux définitions, pour enfin essayer de faire le point sur les éventuelles différences et surtout le statut de ces différences.

Le problème n'est pas tellement de savoir s'il y a des différences. Evidemment, il y en a. Mais c'est de savoir quel statut leur donner.

*

* *

1.1

Prenons comme point de départ les deux textes de 1948. La formule de la quantité d'information que tout le monde connaît se trouve chez Shannon.

[Shannon, 1949, p. 49] « Pouvons-nous définir une quantité qui mesurera en un sens combien d'information est produit par une source d'information discrète ou mieux à quel rythme l'information est produite ? Supposons que nous ayons un ensemble d'événements dont les probabilités d'occurrence sont p_1, p_2, \dots, p_n . Ces probabilités sont connues, mais c'est tout ce que nous savons [...] Pouvons-nous trouver une mesure de la quantité de choix impliqué dans la sélection de l'événement ou du degré d'incertitude quant au résultat ?

[...] Théorème 2 : le seul H satisfaisant les trois postulats précédents est de la forme : $H = -K \sum p_i \log p_i$ où K est une constante positive. »

Cette formule qui donne la mesure de l'information pour le cas discret, dès lors que l'on connaît les probabilités d'apparition des symboles composant le message, ne se trouve à ma connaissance pas chez Wiener. Et pour cause, puisque le cas discret, n'est, dans le livre de 1948, qu'évoqué en quelques pages.

[Cybernetics, p. 61] « Qu'est-ce que l'information et comment la mesurer ? Une des formes les plus simples et les plus unitaires d'information est l'enregistrement d'un choix entre deux alternatives simples également probables, [...] le choix par exemple entre pile ou face lors du lancer d'une pièce. » On retrouve mentionné le logarithme binaire, mais la démonstration dérive tout de suite vers l'étude des ensembles de fonctions continues pour ne retrouver le cas discret qu'à la fin du chapitre, traité en deux ou trois pages. Et dans le traitement du cas discret, Wiener dérive les formules pour la prédiction et le filtrage, mais sans juger bon de produire une mesure de l'information.

Si la mesure de Shannon ne se retrouve pas terme à terme chez Wiener, la dernière partie de l'article de Shannon qui porte sur l'analyse d'une source de messages continue donne la définition de la quantité d'information que l'on retrouve dans Cybernetics.

[Shannon, p. 87] « L'entropie d'un ensemble discret de probabilités p_1, \dots, p_n a été défini comme $H = - \sum p_i \log p_i$. De manière analogue nous définissons l'entropie d'une distribution continue avec la fonction $p(x)$ de densité de distribution par $H = - \int p(x) \log p(x) dx$. » Définition identique à celle donnée dans *Cybernetics* p. 62 comme « une mesure raisonnable de la quantité d'information. »

Qu'est-ce qu'on peut tirer de cette comparaison des mesures ainsi mises bout à bout ?

1.2

Premier point : ces définitions de l'information partagent deux propriétés fondamentales qui en signent l'originalité par rapport aux définitions antérieures que l'on trouvait dans l'ingénierie des communications, dans la tradition des Bell Labs, chez Nyquist et Hartley, et qui sont citées par Shannon au début de son article.

Nous avons donc à faire dans les deux cas à : (a) des définitions probabilistes de la quantité d'information et (b) à des définitions qui intègrent la question du bruit, en fonction de ce canevas probabiliste.

Le point est énoncé de façon tout à fait claire par Shannon au début de son article [p. 31] lorsqu'il présente son travail comme « l'extension de la théorie [de Nyquist et Hartley] pour prendre en compte certains facteurs nouveaux, et en particulier l'effet du bruit sur le canal, et les économies possibles dues à la structure statistique du message original. » Un peu plus loin, évoquant l'usage du logarithme chez Hartley, Shannon ajoute : « La définition [de Hartley] doit être considérablement généralisée pour prendre en compte l'influence des statistiques du message. »

Wiener n'est pas en reste lorsqu'il affirme dès l'introduction du livre de 1948 que [Cybernetics, pp. 8-9] « Un message est une séquence discrète ou continue d'événements mesurables distribué dans le temps, ce que les statisticiens appellent une série temporelle. » Un peu plus loin : « En faisant cela [en améliorant la théorie des filtres] nous avons fait de

l'ingénierie des communications une science statistique, une branche de la mécanique statistique. »

Il faut reconnaître à Wiener que le principe d'une fusion entre le domaine de l'ingénierie des communications et celui de la statistique est déjà énoncé on ne peut plus clairement dans le fameux rapport au NDRC que citait Shannon, dans lequel on lit sous la plume de Wiener, non seulement que la prise en compte des séries temporelles impose une identité d'objet entre ingénierie des communications et statistiques, mais aussi de façon plus précise pour ce qui concerne la mesure de l'information que – je cite - « La transmission d'un élément isolé et fixe d'information ne possède aucune valeur en termes de communication. Nous devons avoir un répertoire des messages possibles et concernant ce répertoire une mesure déterminant la probabilité de ces messages. »

Autrement dit, nous avons ici dès 1942 l'énoncé du principe qui servira à mesurer l'information, sans que cela ne débouche sur la formulation explicite d'une mesure dans le rapport ; puisque l'on n'y trouve, dans l'index, ni mention du terme information, ni mention du terme entropie. Je reviendrai sur la chronologie tout à l'heure.

Pour ce qui est du bruit, la question est évidemment présente chez Wiener aussi dans la mesure où l'un de ses principaux problèmes est celui du filtrage des messages, c'est-à-dire de la restauration autant que faire se peut d'un message qui a été altéré par du bruit. Du coup la question du bruit se trouve intégrée directement à la définition de l'information, puisque la formule de Wiener donne « une quantité d'information infinie pour un message ou une mesure d'une précision absolue, quantité qui devient finie en présence de bruit et qui tend rapidement vers zéro au fur et à mesure que le bruit augmente en intensité. » [Cybernetics, p. 64] Autrement dit, pour obtenir une quantité d'information finie, il faut en prendre en compte l'imprécision de nos mesures ou son équivalent en termes de communication, le bruit.

Si je résumé ce premier point de comparaison. Nous avons deux propriétés fondamentales, la dimension statistique de la théorie, la prise en compte du bruit de ligne.

La 1^{ère} propriété (dimension statistique de la théorie) tient à la formulation même du problème de l'information comme décision au sein d'un ensemble d'événement pour produire une série temporelle. Là-dessus, il semble qu'il y ait un accord fondamental qui autorise à considérer les démarches comme équivalentes.

1.3

Regardons maintenant du côté des différences qui apparaissent sur ce canevas commun.

1.3.1

Nous avons d'abord un ensemble de propriétés différentes entre la mesure pour le cas continu et celle pour le cas discret. Shannon consacre plusieurs pages à l'exposé de neuf différences entre les deux mesures, en insistant en particulier sur le fait que la mesure de l'information dans le cas discret constitue je le cite « une mesure absolue », alors que « la mesure dans le cas continu est relative au système de coordonnées » [Shannon, p. 91], défaut qui n'est pas si important qu'il est paraît puisque les valeurs intéressantes du point de l'ingénieur, le taux de transmission de l'information et la capacité du canal correspondent à des différences entre deux entropies, différence qui est indépendante du système de coordonnées. De plus, la mesure dans le cas continu peut recevoir une valeur négative, ce qui n'empêche pas les taux de transmission et les capacités des canaux d'être toujours positifs.

1.3.2

Mais la question du rapport entre la définition dans le cas continu et dans le cas discret ne se limite pas à la comparaison des propriétés. On voit bien que la principale différence d'accent entre le travail de Wiener et celui de Shannon est que Wiener considère essentiellement le cas continu, là où le travail de Shannon porte essentiellement sur le cas discret.

Je crois que cette proposition est évidente pour Wiener qui à part lors des formulations de principe, lorsqu'il est question de la construction par sélection parmi un ensemble de messages possibles, délaisse le cas discret.

Cela est moins évident pour Shannon, dans la mesure où la théorie mathématique de la communication comporte une discussion relativement détaillée du cas continu. L'article de Shannon comporte en effet 5 parties, l'introduction mise à part, 3 parties portant sur le cas continu, 2 pour le cas discret. Néanmoins, il faut bien noter que le traitement n'est pas égal, dans la mesure où les 2 parties pour le cas discret font à peu près 45 pages, contre seulement 27 pages pour les 3 parties portant sur le cas continu.

Bien plus, Shannon commence par expliquer que – je cite [Shannon, p. 81] « nous n'essayerons pas dans le cas continu d'obtenir les résultats avec la plus grande généralité ou la plus grande rigueur du point de vue des mathématiques pures, dans la mesure où cela impliquerait un long détour par la théorie abstraite de la mesure et obscurcirait la ligne principale de l'analyse. [...] Les libertés prises occasionnellement vis-à-vis de la détermination des limites dans la présente analyse peuvent être justifiées pour tous les cas d'intérêt pratique. »

1.3.3

Je crois que l'on est ainsi justifié à dire que les travaux de Shannon et Wiener occupent une situation symétrique, l'un portant essentiellement sur le cas discret, l'autre sur le continu. Et c'est effectivement dans le développement du cas continu que l'on retrouve chez Shannon l'énoncé de la dette considérable vis-à-vis du travail de Wiener que je citais en introduction.

Reste maintenant une dernière question si on admet cette différence d'accent entre les deux auteurs : comment est-ce que chacun rapporte son domaine de prédilection à celui de l'autre ? Autrement dit, comment est-ce que chacun envisage, à front renversé le rapport du cas discret au cas continu ?

Il y a sur ce point quelque chose qui me semble intéressant.

Si nous regardons du côté de Wiener, le cas discret est considéré sur le plan mathématique comme une pure et simple simplification du cas continu.

[Cybernetics, p. 88-89] « Il existe une autre théorie proche qui en constitue une simplification. Il s'agit de la théorie de la prédiction, des filtres et de la quantité d'information dans les séries temporelles discrètes. [...] La théorie des séries temporelles discrètes est plus simple sous de nombreux aspects que la théorie des séries continues. » Et Wiener donne effectivement une interprétation de ses algorithmes de prédiction et de filtrage pour le cas discret, mais sans donner de formule de la quantité d'information.

Du côté de Shannon, la situation me semble un peu plus complexe. On voit Shannon annoncer que « le cas continu peut être obtenu à partir du cas discret en divisant le continuum des messages et des signaux en un nombre important, mais cependant fini, de petites régions et en calculant les différents paramètres impliqués sur une base discrète. Plus la taille des régions diminue plus la valeur de ces paramètres approche celles du cas continu. [...] Une étude préliminaire indique que la théorie peut être formulée de manière complètement axiomatique et rigoureuse incluant à la fois les cas continus et discrets et bien d'autres. » [Shannon, p. 81]. Autrement dit, on voit Shannon expliquer pour ce que j'en comprends que le cas continu peut être ramené au discret en fonction d'une procédure qui me semble correspondre à une procédure d'échantillonnage. Plus augmente le nombre d'échantillons discrets prélevés sur le signal, plus on se rapproche des valeurs effectives du signal continu.

Nous ne sommes plus dans le principe d'une simplification du continu au discret, mais dans un processus de conversion en quelque sorte du cas continu en cas discret, ce qui permet de réaliser sur un signal continu toutes les opérations permises, en termes de compression et de codage notamment, sur un message discret.

Autrement dit, Wiener comme Shannon a l'air de considérer le domaine de prédilection de l'autre comme un prolongement de son domaine propre. Mais il me semble qu'il y a cependant une dissymétrie, dans la mesure où je ne crois pas que Shannon procède comme il le dit par échantillonnage et à la place on a un travail sur les ensembles de fonction qui ressemble fort à ce que Wiener a développé dans sa théorie des filtres et de la prédiction. Mais il faudrait pouvoir pénétrer plus avant que je ne peux le faire dans l'exposé mathématique pour vérifier ce point.

1.4

Ce qui me semble pouvoir être acquis à l'issue de ce premier parcours en examinant les mesures de l'information telles qu'elles sont énoncées en 1948, c'est qu'il y a manifestement deux voies d'approches de la mesure de l'information, selon que l'on privilégie le discret ou le continu... Ces voies ne sont pas exactement symétriques.

Il est probablement vrai que l'une et l'autre doivent pouvoir être fondue dans une axiomatique plus générale ce que Shannon affirme sans le réaliser ou ce que la démarche de simplification semble montrer chez Wiener.

Reste cependant une différence, me semble-t-il, en dépit de ces passerelles, c'est que chacun des cas de référence se trouve articulé à un ensemble d'applications privilégiées.

Le cas discret chez Shannon, quand bien même il se résumerait à une simplification du cas continu développé par Wiener, a d'abord pour fonction d'aborder un problème qui est celui du codage des messages et disons de la compression de l'information transmise ; problème que Wiener laisse totalement de côté, et lorsqu'il parle de codage, c'est pour en référer à Shannon. Il semble bien que la problématique du codage soit en un sens spécifique au cas discret.

A l'inverse, la question des filtres et encore plus de la prévision des signaux, n'apparaît pas chez Shannon, alors qu'elle constitue le cœur du travail de Wiener.

Du coup, il me semble qu'on peut reconnaître une certaine hétérogénéité sur fond d'homogénéité. Les définitions sont peut-être équivalentes, mais il y a bien un acte, une

décision dans le fait de considérer les deux définitions comme identiques, de dire que codage, filtrage et prévision ont le même objet information.

Après tout, on aurait peut-être pu considérer qu'on avait à faire à deux lignes de travaux parallèles sans considérer qu'il y avait un objet commun : l'information.

Chaque théorie possède après tout son domaine d'objet spécifique, ce qui nous renvoie aux contextes particuliers d'élaboration de ces mesures de l'information.

Je vous propose maintenant de parcourir la chronologie parallèle comparée entre les travaux de Shannon et Wiener pour fixer les contextes dans lesquels ces mesures apparaissent.

2.

Je commence avec le cas de Wiener qui est le plus documenté. Nous possédons plusieurs sources à commencer par les nombreuses relations que Wiener a pu donner de la naissance de la cybernétique, en particulier son autobiographie, *I am a mathematician* en 1956. Mais nous avons aussi la biographie de Masani ou plus récemment le travail pointu de Mindell à qui j'emprunte en grande partie l'exposé suivant.

Automne 1940

Ed Poitras, qui est le conseiller technique de Weaver à la direction D-2 du NDRC, D-2, la section consacrée à la conduite de tir, rencontre Wiener au MIT qui veut appliquer la théorie des communications et des circuits aux problème du contrôle. (Journal de Poitras, cité par Mindell). Wiener s'appuie sur dix ans de travaux avec Yuk Wing Lee, un ancien étudiant chinois du MIT, avec lequel il a travaillé sur l'analyse harmonique et la théorie des réseaux (ou des circuits), avec en projet la construction d'un calculateur analogique, travail qui manquait dira plus tard Wiener d'une compréhension correcte du processus de feedback négatif.

Wiener se tourne alors vers l'application de la théorie des réseaux à la construction d'un dispositif de prédiction qui puisse servir en conduite de tir.

1^{er} décembre 1940

Un contrat est signé avec le NDRC, c'est le projet 6, pour une « théorie mathématique générale de la prédiction et applications ». Wiener embauche à ce moment là un ingénieur Julian Bigelow. Le but est de produire un circuit électrique qui soit capable de suivre une courbe, analogue à une trajectoire d'avion, et d'extrapoler la valeur de la courbe pour une durée déterminée dans le futur.

Un projet minuscule, 3 personnes, au financement infime à comparer au projet parallèle mené aux Bell Labs qui mènera à la construction du canon M9 et sur lequel Shannon sera un temps engagé.

Début de l'année 1941

Wiener et Bigelow construisent une machine pour simuler leurs idées sur la prédiction. Ils rencontrent à ce moment là un problème drastique de fiabilité. Dès que la courbe n'a plus la forme d'une sinusoïde, la machine entre dans des oscillations incontrôlées.

Wiener réalise alors que ce problème est fondamental et ne peut être éliminé. Il est « dans l'ordre des choses ». Il s'agit d'une limite comparable au principe d'incertitude de Heisenberg.

De là, une nouvelle approche apparaît nécessaire, une nouvelle approche statistique. Bigelow et Wiener se lancent alors dans la construction d'un nouveau modèle de predictor fondé sur

une analyse statistique de la corrélation entre la performance passée d'une fonction temporelle et sa performance présente et future. On calcule la position future en fonction des propriétés statistiques de la performance passée, avec révision en continu de la prédiction en fonction des données disponibles en comparant le mouvement de la cible aux prévisions passées.

Juin 1941

Wiener et Bigelow réalisent un filtre électrique pour la prévision qu'ils présentent à l'équipe des Bell Labs. Bode et Lowell sont impressionné, mais insistent sur le fait qu'ils cherchent eux un résultat immédiat, alors que Wiener poursuit un but de plus long terme qui est celui d'une prédiction optimale.

Pendant la fin de l'année

Wiener affine sa théorie statistique au tableau noir, pendant que Bigelow prend des notes. Weaver continue à soutenir le projet persuadé que cette approche permettra de résoudre le problème de l'altitude supposée constante. Nouveau contrat avec la division D-2, Projet 29, pour que Wiener puisse écrire ses résultats.

1^{er} février 1942

« The extrapolation, interpolation and smoothing of stationary time series with engineering applications ». Le rapport au NDRC que mentionne Shannon et qui sera republié en 1949.

Qu'est-ce qu'on trouve dans ce texte capital ? On y trouve une théorie générale du filtrage et de la prédiction pour les séries temporelles, théorie qui n'est pas limitée à la question des trajectoires d'avions, mais inclue les prévisions en économie par exemple.

La contribution principale du livre est de montrer qu'un système à feedback négatif peut optimiser n'importe quelle mesure d'ajustement pour la prévision ou le filtrage.

Est-ce qu'on y trouve une mesure de la quantité d'information ? Non, certainement pas. Le terme d'information n'apparaît même pas dans l'index, pas plus d'ailleurs que le terme d'entropie.

En revanche le terme apparaît bien dans le texte, au tout début, lorsque Wiener expose son projet « une tentative d'unifier la théorie et la pratique de [...] l'analyse des séries temporelles en statistique et de l'ingénierie des communications. »

Un peu plus loin on lit ce qui apparaît comme le principe de cette mesure de l'information, que l'on ne trouvera pas dans le texte : « Le domaine de l'ingénierie des communications est bien plus vaste que celui qu'on lui attribue d'habitude. L'ingénierie des communications porte sur la transmission des messages. Pour l'existence d'un message il est de fait essentiel qu'une information variable soit transmise. La transmission d'un seul élément fixe d'information n'a aucune valeur en termes de communication. Nous devons avoir un répertoire de messages possibles et sur ce répertoire une mesure déterminant la probabilité de ces messages. » Le texte se poursuit, non en développant le principe de la mesure, mais en expliquant que la transmission d'information peut se concevoir aussi bien entre les hommes qu'entre les machines.

Juillet 1942

Wiener et Bigelow font une démonstration de leur predictor à Weaver, Poitras, Fry et Stibitz les responsables de la section conduite de tir, lesquels sont favorablement impressionnés.

Mais le circuit de Wiener-Bigelow rencontre plusieurs problèmes sérieux d'application : d'abord l'algorithme de prédiction suppose une période très longue pour baser sa prévision, ce

qui est rarement réalisé en situation de combat, ensuite l'appareil est extrêmement sensible aux bruits, enfin le critère d'ajustement utilisé n'est pas forcément opératoire pour le tir anti-aérien. S'ajoute à cela évidemment le fait que le circuit de Wiener et Bigelow est beaucoup plus complexe, et donc fragile, que le design plus simple proposé au même moment aux Bell Labs.

Wiener et Bigelow arrivent à la conclusion que l'efficacité de leur predictor est limitée par la connaissance statistique des comportements des pilotes en situation de combat. Ils mettent à ce moment là en place un dispositif de test pour étudier les réactions des opérateurs et collecter des données statistiques. Mais Weaver s'impatience devant l'absence de résultats pratiques et Bigelow se désespère : les comportements des opérateurs humains sont subtils, non linéaires.

Wiener et Bigelow font une tournée des bases américaines pour recueillir des informations. C'est la goûte d'eau qui fait déborder le vase.

Janvier 1943

Weaver met fin au projet au moment de la réorganisation des divisions. C'est l'évènement qui lance la cybernétique, puisque Wiener va à ce moment là chercher un débouché civil à ses recherches militaires. Il se trouve qu'il avait déjà associé son vieil ami Arturo Rosenblueth au printemps 1942.

Janvier 1943

Wiener, Rosenblueth et Bigelow, « Behavior, Purpose and Teleology », dans *Philosophy of Science*, l'article qui lance le programme cybernétique d'alliance avec les sciences du vivant : les comportements finalisés peuvent s'expliquer par des dispositifs mécaniques. Mentionne de manière incidente la conduite de tir, avec l'exemple d'une torpille qui se guide sur sa cible. Mais on n'y trouve à peu près rien concernant l'information.

Voilà pour le contexte des travaux de guerre dans lequel la mesure de l'information trouve son origine.

Il faut attendre la conférence de Wiener à l'Académie des Sciences de New York lors de la session du 21 et 22 octobre 1946 pour voir resurgir la question de la mesure de l'information, dans une conférence qui en quelque sorte le brouillon le canevas de Cybernetics. On retrouvera des formules terme à terme de la conférence dans le livre.

Qu'est-ce qu'on voit apparaître dans cette conférence ? L'inscription de l'ingénierie des communications au sein du contexte plus vaste de la physique statistique, inscription qui se gagne via l'analogie entre information et entropie, qui est exposée ici pour la première, mais sans la mention de l'expérience de pensée du démon de Maxwell, qui elle ne resurgit que dans Cybernetics.

Je cite le passage qui me paraît décisif pour comprendre la manière dont Wiener recontextualise ses recherches de guerre autour de cette analogie information-entropie.

« Nous nous proposons maintenant d'introduire des idées appartenant à la mécanique statistique de Gibbs dans la théorie de la communication. Cette théorie recouvre ce qui est désigné habituellement comme l'ingénierie des communications, mais aussi beaucoup d'autres champs. La théorie du téléphone fait bien sûr partie de l'ingénierie des

communications, mais la théorie des machines à calculer appartient également à ce domaine. De même la théorie des mécanismes de contrôle implique la communication à un organe effecteur et souvent avec un accès en retour, que la machine soit ou non surveillée par un agent humain. Le mécanisme neuromusculaire d'un animal ou d'un homme est certainement un instrument de communication, de même que les organes des sens qui reçoivent des impulsions extérieures. Fondamentalement, les sciences sociales sont l'étude des moyens de communications entre les hommes ou, de façon plus générale, dans n'importe quelle communauté quels que soient les êtres qui la composent. L'idée qui unifie ces diverses disciplines est le message et non quelque appareil particulier agissant sur ces messages. »

Suit une définition du message qui précise les formulations du rapport au NDRC.

« Le message pour transporter de l'information doit représenter un choix entre plusieurs messages possibles. Si tout ce que je peux faire est de créer en bout de ligne un état permanent intégralement caractérisé par son propre passé, alors je cesse de transmettre de l'information. Si j'envoie l'un de ces télégrammes de Noël ou messages d'anniversaires compliqués qu'aiment nos compagnies de téléphone, qui contiennent une quantité importante de verbiage sentimental codé par un nombre entre un et cent, alors la quantité d'information que j'envoie doit être mesurée par le choix entre la centaine d'alternative et n'a rien à voir avec la durée du « message » transcrit.

Il n'est peut-être pas si évident que le message ne présuppose pas seulement un ensemble de messages pour transmettre une signification, mais aussi une distribution a priori sur cet ensemble. [...] Nous voyons ainsi que la notion de message implique les deux idées fondamentales de la mécanique statistique : l'ensemble et la distribution de probabilité sur cet ensemble. Une explication supplémentaire est sans doute requise ici dans la mesure où cela implique aussi l'entropie. L'entropie apparaît ici comme l'inverse de la quantité d'information contenue dans le message. Il n'est peut-être pas évident au premier regard de comprendre pourquoi le logarithme intervient dans la mesure de la quantité d'information. Laissez moi indiquer que l'information de deux sources indépendantes s'additionne alors que leur probabilité sont multipliées, et qu'une variable qui augmente de manière additive alors que l'autre augmente de manière multiplicative est, un facteur constant mis à part, le logarithme de la seconde.

[...] Ainsi la quantité d'information est essentiellement l'inverse de l'entropie. Et cela n'a rien de surprenant qu'information et entropie soit l'inverse l'une de l'autre. L'information mesure l'ordre et l'entropie le désordre. Il est donc possible de concevoir n'importe quel ordre en termes de message. »

Nous n'aurons aucune formulation mathématique, mais les principes de la théorie sont énoncés tels qu'on les trouvera dans les textes de 1948.

Essayons maintenant de retracer la genèse de la définition de Shannon. On a ici beaucoup moins d'information. On n'a pas l'équivalent de l'autobiographie de Wiener, on a une biographie dans les Collected Papers qui reproduisent aussi une interview de 1987 qui donnent plusieurs détails. Et on a quelques autres interviews (Price, Hagenmayer)...

Shannon n'écrivait pas beaucoup. On n'a par exemple aucun brouillon de l'article de 1948.

Quels sont les éléments dont nous disposons ? A mettre en regard de ce que l'on a pu voir du côté de Wiener...

Au moment où les recherches de guerre s'organisent, Shannon est à Princeton avec une bourse postdoctorale.

On connaît très bien ses travaux jusqu'ici.

En 1936 il rejoint le MIT sur un poste d'assistant dans le département d'ingénierie électrique dirigé par Vannevar Bush, qui coordonnera les recherches de guerre. Shannon qui travaille tout en suivant ses études est affecté à l'équipe qui s'occupe de l'analyseur différentiel, le grand calculateur analogique du MIT.

En 1937, alors qu'il travaille pendant l'été aux Bell Labs, il a l'idée d'appliquer l'algèbre de Boole à l'analyse des circuits, comme ceux qui servent au contrôle de l'analyseur différentiel. Il fait de cette analyse le sujet de son master thesis au MIT en 1937-38, Master qui est publié en 1938 et qui remporte un franc succès.

Septembre 1938, Shannon change sur la suggestion de Bush de département pour rejoindre le département de mathématiques.

C'est de cette époque que date le premier document qui mentionne l'intérêt de Shannon pour la théorie de la communication, sous la forme d'une lettre adressé à Bush, le 16 février 1939 dans laquelle Shannon annonce avoir commencé à travailler sur « les propriétés fondamentales des systèmes de transmission de l'information (intelligence) », quel que soit le système, téléphone, radio, télévision, télégraphe. On y trouve une sorte de prémisse du schéma des communications tel qu'on l'aura à partir de 1945, avec un signal qui est une fonction, transformé par un transmetteur, puis reçu et à nouveau transformé par un récepteur. Le problème de la communication dit Shannon est de mesure la distorsion entre f_1 et f_2 . Shannon introduit une mesure de cette distorsion. On peut y voir une sorte de première approche de la question du bruit de ligne, comme dimension incompressible des télécommunications.

En revanche, il n'y a rien qui ressemble à une mesure de la quantité d'information ou même rien qui ressemble à une théorie statistique des communications telle qu'on en a trace à partir de 1943.

Peut-être que l'élément le plus important dans la lettre est la mention de Hartley, dont le nom sera cité au début de l'article de 1948 comme le pionnier avec Nyquist de la mesure de l'information. Shannon explique en février 1939 qu'il va falloir creuser les théorèmes au fondement de l'ingénierie des communications qui n'ont pas encore été explorés à fond. Dans les interviews il dira la même chose : la découverte de la théorie de Hartley a été un choc par la promesse d'une sorte d'approche fondamentale des communications, mais derrière tout restait à prouver.

Manifestement cette perspective n'a pas emballé Bush, qui vient à ce moment là d'être nommé président de la fondation Carnegie à Washington qui comporte un programme consacré à la génétique. Il suggère donc à Shannon d'essayer d'appliquer l'algèbre non plus seulement aux relais, mais à la dynamique des populations. A l'été 1939, Shannon est à Cold Spring Harbor où il arrête le sujet de sa thèse de mathématiques : « une algèbre pour la génétique théorique ».

Printemps 1940, Shannon obtient son diplôme en ingénierie électrique et son PhD en mathématiques.

A l'été 1940, on le retrouve aux Bell Labs où on a deux mémorandums sur l'analyse des circuits.

Durant l'année universitaire 1940-41, Shannon a obtenu une bourse à l'IAS de Princeton pour travailler avec Hermann Weyl sur un sujet de recherche dans le prolongement de sa thèse de génétique théorique. Sujet qui ne l'a jamais vraiment intéressé déclare plus tard Shannon qui affirme poursuivre en sous-main sa théorie de l'information.

Quoi qu'il en soit, il n'a du faire grand-chose, puisqu'en octobre 1940 à la première réunion de la division D-2 du NDRC, le nom de Shannon aurait circulé pour créer un langage standardisé pour les calculateurs mécaniques sur le modèle de sa notation pour l'analyse des circuits.

Le 1^{er} décembre, même jour que Wiener, le contrat 7 est signé avec la D2 pour une étude mathématique des problèmes liés à la conduite de tir, contrat qui court jusqu'au 1^{er} octobre 1941. Et on trouve effectivement d'avril 1941 à juin 1941, 5 rapports au NDRC par Shannon, sous l'étiquette Princeton University.

C'est à la suite de cette période qu'il doit rejoindre à l'invitation de Fry, le directeur du département de mathématiques, les Bell Labs. Mais de fait, dès l'année 1941, il est associé au projet de conduite de tir, puisqu'il participerait dit Segal à la réunion de juin 1941 entre l'équipe des Bell Labs et celle du MIT.

Une fois intégré au département mathématique des Bell Labs on n'a plus de chronologie précise à ma connaissance.

Hagemeyer cite une lettre du 4 juin 1941 adressée à Eisenhart le doyen de Princeton dans laquelle Shannon explique travailler à une théorie générale de la transmission et de la transformation d'information.

C'est assez difficile de dire ce que Shannon a fait exactement à la suite de l'été 1941.

En septembre, on a un court mémo consacré à l'analyse des circuits. En janvier 1942, un long rapport au NDRC sur la théorie et la construction d'une machine à résoudre les équations différentielles linéaires.

Ensuite, plus rien jusqu'au 10 mai 1943 où on a le premier memorandum de Shannon consacré à la cryptographie : « un analogue du système de Vernam pour les séries temporelles continues »

On sait qu'aux Bell Labs les membres du département mathématique servaient de consultants sur divers projets et que Shannon a été employé pour évaluer l'efficacité des systèmes de transmission codée, évaluer l'efficacité des procédures de cryptage et de décryptage d'ailleurs des messages.

Mais il n'a quand même du continuer à toucher à plusieurs puisque l'on retrouve sur la période de mai 1943 à septembre 1945, un mémorandum sur les équations ballistiques, un mémorandum sur les systèmes feedback. Reste qu'entre le mémo de mai 1943 sur le système de Vernam et le fameux rapport de septembre 1945 « A mathematical Theory of Cryptography », la plupart des travaux de Shannon portent sur les systèmes de modulation par impulsion codée, les systèmes PCM. 6 mémorandums.

Si le texte de septembre 1945 est désormais célèbre, c'est que nous y trouvons à peu près tout ce que l'on trouvera dans l'article de 1948. Le rapport de 1945 sera d'ailleurs republié sous une forme remaniée en 1949, sous le titre de « Communication Theory of Secrecy Systems ».

On y retrouve une description générale des communications dans laquelle la clé de chiffrement joue le rôle que tiendra le bruit dans l'article de 1948... et surtout une mesure de l'équivocation ou incertitude qui est identique à la mesure de la quantité d'information de 1948 et qui est fondée sur l'approche statistique employée en cryptographie.

Le cœur du travail se concentre sur l'idée de redondance : un codage parfait est un codage qui élimine la redondance naturelle de la langue, qui est le meilleur outil du cryptanalyste lorsqu'il cherche à décoder un message.

En revanche, j'ignore si le document de 1945 comporte l'analogie entre information et entropie qui est un des points de débat dans la querelle de priorité, puisqu'on sait que Wiener l'introduit en 1946. Segal affirme que l'analogie s'y trouve, mais j'ai aussi lu d'autres sources informées qui disent qu'elle n'y est pas. Comme je n'ai pas le document de 1945, mais seulement la reprise de 1949, je ne sais pas.

Si on revient un peu en arrière, le petit mémorandum de mai 1943 nous donne une indication de l'état d'avancement des travaux de Shannon. Il s'agit d'imaginer un système de cryptage analogue au système de Vernam pour des signaux continus.

On a quelque chose d'intéressant dans la première partie du mémo, là où Shannon montre que le système de Vernam constitue un système parfait, démonstration qu'on retrouvera en 1945, dans la mesure où il maximise l'incertitude. Si on sait qu'un message original est en anglais, alors on sait que la lettre la plus probable est e. Mais avec un système parfait cette information devient inutile dans la mesure où on s'arrange pour que le message encodé soit doté d'une redondance égale.

Il suffit d'appliquer le théorème de Bayes explique Shannon. Autrement dit, on se retrouve face à la première trace d'une application du calcul des probabilités. On voit Shannon reprendre la question en 1945 en comparant les mérites d'une approche fréquentiste et épistémique des probabilités appliquée à la cryptographie.

Si on reprend le dossier du point de vue chronologique. Il est clair que l'on a chez Shannon dès 1945 une formule qui deviendra la formule de la quantité d'information. En revanche l'approche statistique du problème des communications semble présente un peu avant chez Wiener, même si faute de documents pour Shannon c'est difficile de dire quoi que ce soit de définitif. Qui plus est, je ne suis pas sûr qu'on ait affaire au même usage des statistiques dans un cas et dans l'autre.

Reste que ce qui me paraît clair c'est qu'on a à faire à deux lignes de recherches qui sont quand même assez cloisonnées : les problèmes de la cryptographie et du codage ne sont pas immédiatement assimilables au travail sur la conduite de tir et les filtres. Il y a un milieu commun puisque Wiener et Shannon travaillent sur la théorie des circuits, la conduite de tir, mais les lignes divergent avec le travail de Shannon sur la cryptographie qui n'a pas d'équivalent chez Wiener et qui repose fondamentalement sur la manipulation de l'information discrète.

3.

Je pense qu'on peut maintenant mieux comprendre, pour conclure, le statut des différences.

3.1

Commençons par regarder ce qui est dit par l'un et par l'autre du travail de l'autre. On a quelques appréciations divergentes qui sont souvent liées à la querelle de priorité, et qui sont intéressantes dans la mesure où elles nous livrent des indications concernant les différences d'approche.

Je pense qu'on peut maintenant comprendre un peu mieux les termes de la petite querelle de priorité qui se fait jour après coup entre Shannon et Wiener.

Le dossier des attaques est plus fourni chez Wiener, qui s'estime lésé que chez Shannon.

Chez Shannon, on trouve essentiellement, dans une interview tardive de 1987, une petite phrase souvent citée : au journaliste qui l'interroge sur le rôle que Wiener aurait pu jouer sur la réflexion menée aux Bell Labs en affirmant dès 1947 que l'information est de l'entropie, Shannon répond de manière assez sèche : « Je n'avais jamais entendu parler de cette remarque [sur information = entropie] quand j'ai commencé mon travail. Je ne pense pas que Wiener ait grand-chose à voir avec la théorie de l'information. Il n'a pas eu une grande influence pour mes idées, bien que j'ai une fois suivi un de ses cours. Ne vous méprenez pas, c'était un grand mathématicien. C'était une de mes idoles quand j'étais jeune étudiant au MIT. »

Du côté de Wiener, le dossier est plus soutenu.

Nous avons d'abord dans l'entourage de Wiener, chez Bateson et Levinson, plusieurs témoignages qui racontent les visites régulières de Shannon à Wiener pendant la guerre, visites auxquelles Wiener aurait mis fin, ayant l'impression de se faire voler ses idées.

De Wiener lui-même, on trouve deux formules intéressantes.

Je pense d'abord à un passage dans *I am a mathematician*, dans lequel Wiener réinvestit la différence entre le continu et le discret : [I am, p. 263].

« J'ai approché la théorie de l'information à partir d'un circuit électrique à courant continu ou du moins de quelque chose qui pouvait s'interpréter comme un courant continu. Au même moment, Claude Shannon des Bell Telephone Laboratories développait une théorie parallèle et largement équivalente du point de vue de système de relai électriques. Ils représentent un développement direct à partir de son travail précédent sur l'utilisation de la logique dans les problèmes de relais. Comme je l'ai déjà dit auparavant, Shannon adore le discret et évite le continu. Il a considéré les messages discrets à la manière d'une séquence de oui et de non distribués dans le temps et a regardé les décisions simples entre un oui et un non comme l'élément d'information. Dans la théorie continue des filtres, j'ai été amené à une définition très proche de l'unité d'information, à partir de ce qui était au point de départ un point de vue très différent. En introduisant la définition de la quantité d'information de Shannon-Wiener (car elle appartient à part égale à chacun de nous deux), nous avons rompu de manière radicale avec l'état de l'art existant sur le sujet... »

On retrouve cette idée qui me semble assez juste de deux définitions équivalentes, mais qui pourtant s'énoncent à partir de points de vue très différents.

Une dernière citation pour finir, celle-ci beaucoup plus désagréable, qui date de 1953, dans un texte non publié, puisqu'il s'agit d'une lettre de Wiener à Francis Bello, le rédacteur en chef de la section technique du magazine fortune qui désirait écrire un article sur la théorie de l'information et ses inventeurs.

« Le Dr. Shannon est un employé de la compagnie Bell, écrit Wiener, et est amené à développer les notions de communication dans un registre limité qui correspond aux intérêts de sa compagnie. Il doit travailler beaucoup plus en fonction de résultats immédiatement exploitables que je ne suis tenu de le faire. Je suis professeur d'université et j'ai découvert que le nouveau royaume des idées de la communication constitue une source fertile de nouveaux concepts, non seulement pour la théorie de la communication, mais aussi pour l'étude de l'organisme vivant et de nombreux autres problèmes connexes. » [Dark Hero, n. p. 375]. Voici ce qu'on peut appeler la sociologie spontanée des savants.

Ce qui me frappe c'est que ces déclarations ne me paraissent pas inconciliables. Si au final, les travaux se rejoignent et que d'une manière ou d'une autre, par échantillonnage, par simplification, ou dans l'horizon d'une axiomatique à venir, le développement de la mesure de l'information dans le cas continu et celle dans le cas discret se rejoignent, on a à faire à des lignes qui sont largement parallèles et qui obéissent à des logiques et des démarches qui me semblent assez différentes.

Ce n'est pas insensé de la part de Shannon de dire que Wiener n'a pas grand-chose à voir dans la théorie de l'information entendue comme théorie du codage. On trouve un point symétrique dans la déclaration de 1956 chez Wiener.

Autrement dit, peut-être que la mesure est la même, au final, mais ce qui compte c'est aussi ce qu'on en fait. Et posséder la mesure de Wiener n'autorise pas à résoudre de manière automatique le problème du codage posé par Shannon... et vice et versa, posséder la mesure de Shannon ne nous donne pas en claquant des doigts un algorithme de prédiction optimal. Il faut distinguer la mesure et ce que l'on en fait.

Pour ce qui est de la lettre de 1953, ce que dit Wiener correspond à ce qui s'est passé, c'est-à-dire que Wiener a manifestement cherché à étendre la théorie de l'information au-delà du cas strict de l'ingénierie des communications, ce que effectivement Shannon, à tort ou à raison, a toujours été réticent à faire.

3.2

La différence des projets se marque par le statut et le mode d'écriture des textes.

Shannon publie dans le Bell System Technical Journal, Wiener publie son livre dans la collection des actualités scientifiques chez Hermann, qui associe toujours résultats scientifiques et philosophie des sciences.

Plus encore, la théorie des communications chez Wiener n'apparaît que comme un chapitre dans un projet plus vaste qui consiste à étendre le royaume de l'information du domaine des communications jusqu'à l'étude des organismes vivants voir des sociétés, en passant par les domaines du calcul mécanique ou du rétrocontrôle.

Cette différence dans le projet se retrouve dans les textes. Shannon donne une sorte d'axiomatique des communications qui est en quelque sorte fermée sur elle-même et ses applications là où Wiener cherche à inclure la théorie des communications comme un cas particulier dans une approche plus vaste, qui est une physique statistique de l'information.

Je crois que la différence se ressent par exemple pour ce qui est de l'usage des probabilités. Chez Shannon, nous avons une application stricte du calcul des probabilités aux problèmes du codage, qui découle sans doute des travaux en cryptographie.

Chez Wiener la dimension statistique apparaît toujours en lien avec l'arrière-plan physicaliste de la théorie. Si Wiener utilise les outils statistiques c'est toujours en référence à sa théorie du mouvement brownien ; les séries temporelles qu'étudient la théorie des communications n'apparaissent plus que comme un cas particulier de ces séries statistiques qui s'observent dans un univers aléatoire et contingent, si on lit Wiener.

3.3

Je crois que du coup on retrouve en dépit de l'identité entre les formules des représentations assez distinctes de ce que peut être la quantité d'information, au croisement entre deux paramètres :

- la distinction entre le cas continu et le cas discret, l'analogique et le numérique
- mais aussi le statut donné à la théorie de la communication qui conduit Wiener à n'en faire qu'un chapitre dans une théorie plus générale qui est une physique statistique de l'information.

Ces deux différences : une axiomatique du code chez Shannon contre une physique statistique de l'information chez Wiener, avec priorité donnée au signal continu sur le symbole discret, sont manifestement ancrées dans les contextes de production de ces deux mesures de l'information.

L'effet du contexte est évident pour ce qui est des domaines d'origine et des applications privilégiées dont j'ai retracé la chronologie parallèle, mais il me semble aussi que cela dépend des conceptions de la science que se font Wiener et Shannon, avec dans un cas une théorie qui vaut par ses applications extraordinaires dans le domaine du code, dans l'autre une théorie qui s'inscrit dans une démarche plus vaste, y compris avec une composante philosophique, sur le développement de la mécanique statistique.

Ces attentes divergentes vis-à-vis de ce que peut être une théorie de l'information sont sans doute liées au statut des deux protagonistes, l'un professeur de mathématiques au MIT, l'autre chercheur au sein des Laboratoires Bell, comme le disait Wiener en 1953, en laissant de côté la dimension péjorative associée à la recherche industrielle dans le texte de Wiener.

Voilà donc qui au final me semble valider l'idée d'une représentation non unifiée, clivée, de l'information, telle que j'ai pu la proposer ailleurs, qui derrière l'unité de la formule s'interroge sur ce qui est visé sous le nom d'information.

Je vous remercie.