ANNALES DE L'I. H. P., SECTION A

MIOARA MUGUR-SCHÄCHTER

Le concept nouveau de fonctionnelle d'opacité d'une statistique. Étude des relations entre la loi des grands nombres, l'entropie informationnelle et l'entropie statistique

Annales de l'I. H. P., section A, tome 32, nº 1 (1980), p. 33-71

<http://www.numdam.org/item?id=AIHPA_1980__32_1_33_0>

Ann. Inst. Henri Poincaré,
Vol. XXXII, n° 1, 1980, p. 33-71.

Section A :
Physique théorique.

Le concept nouveau de fonctionnelle d'opacité d'une statistique. Étude des relations entre la loi des grands nombres, l'entropie informationnelle et l'entropie statistique.

par

Mioara MUGUR-SCHÄCHTER
Université de Reims, Laboratoire de Mécanique Quantique

RÉSUMÉ. — Un nouveau concept probabiliste — l'opacité d'une statistique vis-à-vis de la loi de probabilité sous-jacente — est construit à l'intérieur de la théorie abstraite des probabilités. Ce concept montre que l'entropie abstraite d'une statistique évolutive, l'entropie « informationnelle » ou « probabiliste » constante de la loi de probabilité sous-jacente, et la loi des grands nombres, sont des aspects différents et emplacés sur des niveaux distincts d'organisation logique, d'une entité descriptive unique qui fournit une mesure de la signifiance de la statistique considérée. Les perspectives ouvertes par les résultats établis sont brièvement définies. Elles établissent une base formelle détaillée pour l'analyse de la signification des concepts d'entropie physique.

I. INTRODUCTION

Ce travail s'inscrit dans le projet d'élucider en profondeur les relations entre l'entropie statistique physique et l'entropie informationnelle.

L'entropie $S = -k \sum_i n_i \log n_i$ définie par Boltzmann pour un ensemble statistique de systèmes physiques joue depuis plus de 100 ans un rôle fondamental dans les descriptions théoriques de la réalité physique. Ce concept exerce une sollicitation obsédante sur ceux qui abordent les problèmes de l'irréversibilité, du temps, des évolutions organisatrices du vivant.

D'autre part, l'entropie informationnelle $H = -\sum_i p_i \log p_i$ définie par Shannon en 1948 joue, de son côté, un rôle de plus en plus remarquable dans les descriptions des processus qui transmettent des *morphismes* — que l'on appelle uniformément des « informations » — mais qui peuvent être aussi bien matériels, comme ceux que véhiculent les phénomènes génétiques, qu'abstraits, comme ceux portés par des langages. Ces descriptions suggèrent en particulier la possibilité d'une véritable théorie mathématique des évolutions organisatrices.

Mais au-delà, l'on sent s'ébaucher les traits d'une grande synthèse entre les descriptions des divers aspects du réel et la description des connexions du réel avec l'esprit. Les relations profondes mais tellement obscures encore entre les théories et l'épistémologie sont en cours de se structurer dans une formalisation unifiante.

Ces deux concepts d'entropie, celui de Boltzmann et celui de Shannon, définis ainsi tout à fait indépendamment l'un de l'autre, ont une similitude formelle évidente, que Shannon a soulignée en rechoisissant le mot entropie pour dénommer sa fonction H. Mais les contenus sémantiques de ces deux concepts, ainsi que la relation entre ces contenus sémantiques, se sont à ce jour dérobés à une véritable compréhension. Les quelques considérations bien connues de L. Brillouin à ce sujet sont loin d'avoir épuisé la question.

Selon certains auteurs cette relation sémantique serait très faible, parce que le concept de Boltzmann, étant « lié » dimensionnellement à une énergie par le facteur k est foncièrement physique, et qu'en outre il caractérise des *évolutions*, tandis que le concept de Shannon est « libre » dimensionnellement (c'est un nombre) et il est significatif pour des *transmissions* (ou « communications ») de morphismes mais *pas* pour des évolutions, car sa structure et sa valeur restent *constantes* à l'intérieur de tout schéma probabiliste où il a été défini. Selon d'autres auteurs, au contraire, il existerait entre les significations de l'entropie probabiliste H et de l'entropie statistique S, une relation profonde qui se reflète dans la similitude de leurs expressions formelles. Cette controverse est importante. L'entropie de Boltzmann et celle de Shannon ont manifesté chacune une grande puissance descriptive. Ceci indique que l'une comme l'autre elles ont capté certains traits essentiels du réel. Mais il arrive parfois dans les progressions de la Science que l'intuition sémantique qui engendre un concept descriptif efficace devance longuement la connaissance explicite et exacte du contenu sémantique dont a été doté ce concept. Cependant seule une telle connaissance explicite peut permettre d'utiliser exhaustivement les ressources descriptives du concept, sans s'empêtrer dans de fausses limitations ou « problèmes », ni glisser dans des extensions inconsistantes avec les inévitables limitations effectivement introduites au départ. Ceci justifie le but d'expliciter les contenus sémantiques des deux concepts d'entropie, ainsi que leurs relations.

Mais ce but est entaché d'une certaine inhomogénéité qui le rend inaccessible *directement* : l'entropie $H = -\sum_i p_i \log p_i$ de Shannon est un concept probabiliste *pur*, abstrait, que l'on est libre d'insérer dans les applications interprétatives les plus diverses *via* des hypothèses interprétatives adéquates; tandis nous l'entropie de Boltzmann $S = -k\sum_i n_i \log n_i$ est un concept probabiliste intégré d'emblée dans la description d'une classe de situations physiques. Or, deux concepts qui s'emplacent sur des niveaux de généralité différents, quelles que soient leurs ressemblances, ne sont pas directement comparables. Nous avons alors choisi l'attitude suivante :

Dans le travail présent nous nous plaçons préliminairement à l'intérieur de la théorie *abstraite* des probabilités. Toute connexion sémantique des concepts abstraits définis ici, avec telle ou telle entité réelle, est remise à des étapes ultérieures. Dans ce cadre purement formel, une démarche tout à fait naturelle et simple conduit ([*1*]) à une certaine fonctionnelle, « la fonctionnelle d'opacité d'une statistique par rapport à la loi de probabilité sous-jacente ». Cette fonctionnelle est liée à la valeur de la « métaprobabilité » de réalisation de la statistique considérée, à la limite des grands nombres; en même temps elle mesure l'écart de la statistique considérée, par rapport à la statistique qui exprimerait le plus rigoureusement la loi de probabilité sous-jacente.

D'autres concepts probabilistes déjà construits précédemment mesurent eux aussi cet écart, plus ou moins explicitement (par exemple la vraisemblance). Mais la fonctionnelle d'opacité construite ici introduit ce trait nouveau qu'elle établit en plus une synthèse conceptuelle entre l'entropie probabiliste de Shannon, un certain concept *abstrait* d' « entropie statistique » (qui paraît avoir des liens étroits avec le concept d'entropie physique de Boltzmann) et la loi des grands nombres.

Cette synthèse, obtenue au niveau abstrait, crée diverses perspectives. Elle fournit une base appropriée pour l'étude des liens existant entre l'entropie probabiliste de Shannon et les concepts d'entropies statistiques physiques. Elle nous semble ouvrir ainsi la possibilité de réaborder à partir du niveau conceptuel le plus général, l'unification que l'on tente depuis quelque temps entre les approches de la Thermodynamique Statistique et les approches informationnelles (Jaynes [*2*], Tribus [*3*], Katz [*4*], etc.). Les significations véritables des évolutions d'entropies physiques, du microcosmos et jusque dans les cosmologies, pourront être peut-être comprises enfin plus clairement. En outre il devient concevable de tenter d'étendre la Thermodynamique Statistique en dessous de la phénoménologie des processus irréversibles développée par Prigogine [5] et par son école, en élaborant une théorie *probabiliste* de ces processus. Très généralement, la synthèse conceptuelle accomplie ici au niveau purement abstrait nous semble devoir toucher dans ses conséquences toutes les descriptions statistiques d'évolutions réelles, qu'elles soient « organisatrices » ou « désorganisatrices », qu'il s'agisse de substance physique (anorganique ou vivante), de comportements (individuels ou collectifs), ou même des progressions de la connaissance. Enfin, le problème épistémologique des rapports entre la conceptualisation individuelle et la conceptualisation probabiliste, et du rôle particulier que tient en ces conceptualisations ce que l'on appelle « le temps », semble devenir abordable en termes mathématiques.

Mais, bien sûr, ce ne sont pour l'instant rien de plus que des conjectures. Seules des explorations futures effectives pourront déterminer la force descriptive que le nouveau concept abstrait d'opacité possède véritablement.

II. ANNONCE MÉTHODOLOGIQUE

Le développement qui suit obéit à la méthodologie russelienne de la hiérarchisation des types logiques. A ce jour, cette méthodologie reste encore très rarement utilisée en physique et même en mathématique, bien qu'en logique elle soit maintenant devenue usuelle. Cette méthologie tient compte systématiquement d'une caractéristique qui est inhérente à toute description ensembliste, à savoir : étant choisies certaines entités de départ ε_0, les ensembles $\varepsilon_1 = \{ \varepsilon_0 \}$ de telles entités introduisent des propriétés spécifiques, i. e. qui n'existent pas pour les entités ε_0 et qui ne sont pas définissables pour les ensembles $\varepsilon_2 = \{ \varepsilon_1 \} = \{ \{ \varepsilon_0 \} \}$ d'ensembles $\varepsilon_1 = \{ \varepsilon_0 \}$ d'entités ε_0. Plus généralement, chaque niveau nouveau n d'organisation ensembliste des entités ε_0 en ensembles $\varepsilon_n = \{ \varepsilon_{n-1} \} = \{ \{ \varepsilon_{n-2} \} \} = \dots$ introduit certaines propriétés spécifiques de ce niveau n, qui ne sont définissables ni pour les ensembles ε_{n-1} du niveau précédent, ni pour les ensembles ε_{n+1} du niveau supérieur. Pourtant les propriétés spécifiques du niveau n sont reliées aux propriétés spécifiques du niveau $n - 1$ et à celles du niveau $n + 1$ d'une manière qui dépend de ε_0 et de n. En ce sens toute description ensembliste est marquée par un caractère « d'organisation hiérarchisée ». En conséquence de ce caractère il est évident que les traitements logico-mathématiques de niveaux d'organisation ensembliste différents doivent être clairement distingués les uns des autres. C'est une condition *sine qua non* si l'on veut d'une part éviter tous faux problèmes (paradoxes) et, d'autre part, si l'on veut assurer une descriptibilité à la fois rigoureuse et pragmatiquement efficace des propriétés spécifiques de chaque niveau et des relations qui existent entre des propriétés différentes d'un même niveau ou entre des propriétés de niveaux différents.

Les biologistes, les sociologues, les linguistes, les philosophes des sciences, les penseurs interdisciplinaires, deviennent de plus en plus conscients du caractère essentiellement hiérarchique de toute « organisation » et de la grande importance méthodologique d'une prise en considération systématique de ce caractère. Cette évolution de la pensée se manifeste clairement dans le courant structuraliste et surtout dans les oppositions bien connues entre les attitudes réductionnistes et les attitudes holistes. Pourtant la conscience du fait que la structure hiérarchique de toute organisation *présuppose* une ségrégation ensembliste de l'objet d'étude, nous semble rester encore confuse. Les diverses descriptions existantes d'organisations hiérarchiques ne se rattachent pas explicitement, à notre connaissance, à la théorie ensembliste de Russel, des types logiques hiérarchisés. D'autre part, les conséquences méthodologiques qui découlent du caractère hiérarchique de toute organisation semblent n'avoir influencé que faiblement les démarches des physiciens et même celles des mathématiciens. C'est peut-être à cause de la composition de ces deux zones d'ombre, que nous n'avons jamais rencontré une illustration délibérée de l'efficacité méthodologique de la distinction des types logiques, qui soit accomplie à l'aide d'un développement mathématique susceptible d'applications ultérieures à des descriptions du réel. Dans ce qui suit l'on pourra trouver précisément une telle illustration.

III. DÉFINITIONS PRÉALABLES

Nous allons tout d'abord préciser quelques éléments de langage qui nous sont plus ou moins propres. Pour cela nous devons revenir — sans détails inutiles — sur les concepts courants de phénomène aléatoire et d'espace de probabilité.

Souvent l'on peut spécifier une certaine procédure $\mathscr{P}$ qui est posée comme étant « identiquement » réitérable un nombre arbitraire de fois (1) et qui, lors de chacune de ses réitérations, comporte comme « effet » un certain « événement élémentaire » qui en général varie d'une réitération de $\mathscr{P}$ à une autre, nonobstant l'identité supposée de toutes ces réitérations. La paire formée de la procédure identiquement réitérable $\mathscr{P}$ et de « l'univers » $\mathscr{U} = \{\varepsilon_i\}$ de tous les effets possibles de $\mathscr{P}$, constitue un « phénomène aléatoire » $(\mathscr{P}, \mathscr{U})$. Dans tout ce qui suit l'univers $\mathscr{U} = \{\varepsilon_i\}$ est supposé *fini* : $i = 1, 2, \ldots$.

A partir de l'univers $\mathscr{U}$ d'événements élémentaires comportés par un

(1) Il n'a jamais été analysé — et ce serait important — à *quoi* exactement s'applique la qualification « identiquement ».

phénomène aléatoire, l'on peut construire divers « espaces de probabilité » $(\mathscr{U}, \tau, p(\tau))$, en définissant sur $\mathscr{U}$ une certaine tribu τ d' « événements », i. e. de parties mesurables de $\mathscr{U}$, et en posant sur cette tribu une certaine mesure de probabilité $p(\tau)$. Soit $(\mathscr{U}, \tau_t, p(\tau_t))$ *l'espace de probabilité fondamental sur* $(\mathscr{P}, \mathscr{U})$, où τ_t est la tribu *totale* sur $\mathscr{U}$, qui en particulier contient chacun des $\varepsilon_i \in \mathscr{U}$. L'ensemble $\{ p_i, \forall_i \}$ des probabilités que la mesure $p(\tau_t)$ assigne aux événements élémentaires $\varepsilon_i \in \mathscr{U}$, sera appelé *la loi de probabilité fondamentale de* $(\mathscr{P}, \mathscr{U})$.

Mais il est également possible de définir à partir de $(\mathscr{P}, \mathscr{U})$ une autre sorte d'espaces de probabilité, en un certain sens dérivés : soit le phénomène aléatoire $(\mathscr{P}^{\mathrm{N}}, \mathscr{U}^{\mathrm{N}})$ où l'univers $\mathscr{U}^{\mathrm{N}} = \{ \sigma_x^{\mathrm{N}}, \forall x \}$ est l'ensemble de toutes les suites $\sigma_x^{\mathrm{N}} = (\varepsilon_i^{(1)}, \dots \varepsilon_s^{(l)}, \dots \varepsilon_k^{(\mathrm{N})})$ possibles de N événements $\varepsilon_i \in \mathscr{U}$, tout indice i pouvant se répéter un nombre de fois variant de 1 à N, et chaque telle suite étant obtenue comme un effet possible de la procédure $\mathscr{P}^{\mathrm{N}} = \mathscr{P}^{(1)}.\mathscr{P}^{(2)} \dots \mathscr{P}^{(\mathrm{N})}$ constituée par N réitérations identiques de $\mathscr{P}$, en succession ou en simultanéité (l'indice supérieur $(l) = (1), (2) \dots (\mathrm{N})$ qui dans une suite $\sigma_x^{\mathrm{N}} = (\varepsilon_i^{(1)} \dots \varepsilon_s^{(l)} \dots \varepsilon_k^{(\mathrm{N})})$ affecte un événement $\varepsilon_s^{(l)}$, étiquette celle des réalisations de $\mathscr{P}$ qui a amené cet ε_s; l'indice global x dans le symbole σ_x^{N} singularise l'entière structure de la suite, *compte tenu* de la correspondance $\mathscr{P}^{(l)} \leftrightarrow \varepsilon_s^{(l)}$ i. e. de la correspondance des indices d'ordre $(l) = (1), (2) \dots (\mathrm{N})$ des $\mathscr{P}^{(l)}$ avec les indices d'événement $i = 1, 2 \dots \lambda$ des ε_i. Une telle suite σ_x^{N} maximalement étiquetée sera dénommée une complexion individuelle de puissance N. L'espace de probabilité $(\mathscr{U}^{\mathrm{N}}, \tau_t^{\mathrm{N}}, p_{\mathrm{N}}(\tau_t^{\mathrm{N}}))$ où τ_t^{N} est la tribu *totale* sur $\mathscr{U}^{\mathrm{N}}$, sera dénommé *métaespace de puissance N basé sur* $(\mathscr{P}, \mathscr{U})$. En particulier τ_t^{N} contient toutes les complexions individuelles de puissance N, σ_x^{N}, possibles, donc $p_{\mathrm{N}}(\tau_t^{\mathrm{N}})$ spécifie pour chaque telle complexion individuelle une certaine probabilité $p_{\mathrm{N}}(\sigma_x^{\mathrm{N}})$. Nous posons que

$$p_{\mathrm{N}}(\sigma_x^{\mathrm{N}}) = p(\varepsilon_i^{(1)}) \dots p(\varepsilon_s^{(l)}) \dots p(\varepsilon_k^{(\mathrm{N})}) = \prod_{i=1}^{\lambda} p_i^{n_i} \tag{1}$$

où $p_i = p(\varepsilon_i)$ et l'exposant n_i, $0 \leq n_i \leq \mathrm{N}$, $\sum_{i=1}^{\lambda} n_i = \mathrm{N}$, est le nombre des apparitions en σ_x^{N} de ε_i. L'hypothèse (1) est conçue ici comme *une caractéristique de définition de ce que nous appelons un phénomène aléatoire* $(\mathscr{P}^{\mathrm{N}}, \mathscr{U}^{\mathrm{N}})$ (à la différence d'une « source » au sens de la Théorie de l'Information) : « l'identité » posée pour les N réitérations de $\mathscr{P}$ dans $\mathscr{P}^{\mathrm{N}} = \mathscr{P}^{(1)} \dots \mathscr{P}^{(l)} \dots \mathscr{P}^{(\mathrm{N})}$, qui exige de revenir aux *mêmes conditions* pour chacune de ces réitérations, tranche $\mathscr{P}^{\mathrm{N}}$ en N répétitions « indépendantes » de $\mathscr{P}$ (cette exigence maintient dans la classe des sources à mémoire nulle).

La loi de probabilité fondamentale $\{p_i\}$ concerne les événements élémentaires ε_i qui constituent un niveau d'organisation de départ. La loi de probabilité $p_N(\sigma_x^N)$ de (1) concerne un second niveau d'organisation où les événements considérés sont les ensembles maximalement étiquetés $\sigma_x^N = (\varepsilon_1^{(1)} \dots \varepsilon_i^{(l)} \dots \varepsilon_k^{(N)})$ de N événements $\varepsilon_i^{(l)}$. Si l'on veut maintenant caractériser la seule structure *statistique* d'une suite $\sigma_x^N \in \mathscr{U}^N$, i. e. abstraction faite des indices supérieurs d'ordre, donc abstraction faite de la correspondance $\mathscr{P}^{(l)} \leftrightarrow \varepsilon_i^{(l)}$ l'on est conduit sur un troisième niveau d'organisation :

Soient $0 \leq \left(\frac{n_i}{N}\right)_x \leq 1$, $i = 1, 2 \dots \lambda$, $\sum_{i=1}^{\lambda} \left(\frac{n_i}{N}\right)_x = 1$ les λ fréquences relatives d'apparition dans σ_x^N des événements ε_i. L'ensemble $\left\{ \left(\frac{n_i}{N}\right)_x \right\}_j$ de celles-ci, que l'on peut caractériser globalement par un indice unique j, définit entièrement la structure statistique de σ_x^N. Mais la *même* structure statistique j peut également se réaliser pour d'autres suites $\sigma_{x'}^N \in \mathscr{U}^N$ ou $\sigma_x^{N'} \in \mathscr{U}'$ ayant soit une autre structure des correspondances $\mathscr{P}^{(l)} \leftrightarrow \varepsilon_s^{(l)}$ appelant un indice global $x' \neq x$, soit une autre puissance $N' \neq N$ mettant en jeu un autre univers $\mathscr{U}^{N'} \neq \mathscr{U}^N$. En ce sens la structure statistique j fixée par un ensemble donné de λ rapports $\frac{n_i}{N}$, est une propriété qui transgresse le niveau caractéristique des complexions individuelles σ_x^N (bien que chaque σ_x^N donnée possède cette propriété) car elle caractérise certains ensembles de σ_x^N à l'intérieur desquels x et N peuvent varier. Cette propriété requiert donc une définition et une dénomination propres.

Par définition nous appelons *complexion statistique* $\mathscr{C}_j$ (ou j) un ensemble

$$\left\{ \left(\frac{n_1}{N}\right)_j, \dots, \left(\frac{n_i}{N}\right)_j, \dots, \left(\frac{n_\lambda}{N}\right)_j \right\} = \{ f_{1j}, \dots, f_{ij}, \dots, f_{\lambda j} \}$$

de λ rapports

$$\left(\frac{n_i}{N}\right)_j = f_{ij}, \qquad 0 \leq f_{ij} \leq 1, \qquad \sum_{i=1}^{\lambda} f_{ij} = 1,$$

caractérisé *globalement* par l'indice unique $j = 1, 2 \dots M(\lambda, N)$ qui indique l'un des $M(\lambda, N) = \frac{(N + \lambda - 1)!}{(\lambda - 1)!\,N!}$ tels ensembles possibles de λ rapports $\left(\frac{n_i}{N}\right)_j$. N étant fixé, *pas* toutes les complexions $\mathscr{C}_j = \{ f_{ij}, \forall_i \}$ concevables sur l'univers $\mathscr{U}^N = \{ \varepsilon_i \}$, $i = 1, 2 \dots \lambda$ peuvent se réaliser *dans* $(\mathscr{U}^N, \tau_t^N, p(\tau_t^N))$, car les rapports possibles $\frac{n_i}{N}$ dépendent de N.

Il est important de remarquer que, à la différence du concept de complexion statistique utilisé en thermodynamique statistique, où il est tenu compte des valeurs *absolues* des populations d'événements élémentaires, notre concept de complexion statistique ne tient compte que des populations *relatives* $\frac{n_i}{\mathrm{N}}$.

IV. CONSTRUCTION D'UNE FONCTIONNELLE D'OPACITÉ D'UNE COMPLEXION STATISTIQUE PAR RAPPORT A LA LOI DE PROBABILITÉ FONDAMENTALE

Estimons maintenant la probabilité de réalisation — dans la métaespace $(\mathscr{U}^{\mathrm{N}}, \tau_t^{\mathrm{N}}, p_{\mathrm{N}}(\tau_t^{\mathrm{N}}))$ — d'une suite σ_x^{N} ayant une structure statistique j donnée.

Toute suite $\sigma_x^{\mathrm{N}} \in \mathscr{U}^{\mathrm{N}}$ possède nécessairement une certaine structure statistique, et plusieurs suites σ_x^{N} qui diffèrent par les correspondances $\mathscr{P}^{(l)} \leftrightarrow \varepsilon_s^{(l)}$ peuvent avoir la même statistique j. L'on peut donc former une partition $\tau_t^{\mathrm{N}} = \bigcup_{j=1}^{\mathrm{M}(\lambda,\mathrm{N})} \{ \sigma_x^{\mathrm{N}} : \mathscr{C}(\sigma_x^{\mathrm{N}}) = \mathscr{C}_j \}$ de τ_t^{N} où $\{ \sigma_x^{\mathrm{N}} : \mathscr{C}(\sigma_x^{\mathrm{N}}) = \mathscr{C}_j \} = \{ \sigma_x^{\mathrm{N}} \}_j$ est l'ensemble de toutes les suites $\sigma_x^{\mathrm{N}} \in \mathscr{U}^{\mathrm{N}}$ ayant la même complexion statistique j mais se distinguant l'une de l'autre par l'indice x. La mesure de probabilité $p_{\mathrm{N}}(\tau_t^{\mathrm{N}})$ assigne à chaque événement $\{ \sigma_x^{\mathrm{N}} \}_j \in \tau_t^{\mathrm{N}}$ une certaine probabilité $p_{\mathrm{N}}(\{ \sigma_x^{\mathrm{N}} \}_j) = p_{\mathrm{N}}(\mathscr{C}_j)$ qui par rapport à la probabilité $p_{\mathrm{N}}(\sigma_x^{\mathrm{N}})$ de (1) s'emplace à un niveau d'organisation supérieur. Cette probabilité est déterminée par (1) et par la loi des probabilités totales. L'on obtient évidemment

$$p_{\mathrm{N}}(\mathscr{C}_j) = \frac{\mathrm{N}\,!}{\prod_{i=1}^{\lambda} n_i(j, \mathrm{N})\,!} \prod_{i=1}^{\lambda} p_i^{n_i(j,\mathrm{N})} \tag{2}$$

où $n_i(j, \mathrm{N}) = n_i$ dénote le numérateur de la fraction $\left(\frac{n_i}{\mathrm{N}}\right)_j = f_{ij}$ et où le facteur $\mathrm{N}\,!/\Pi_i n_i\,!$ est la permutabilité d'une séquence σ_x^{N} ayant la complexion j. Formons maintenant la quantité $-\log p_{\mathrm{N}}(\mathscr{C}_j)$. En tenant compte de la formule de Stirling $n\,! = \sqrt{2\pi} n^{(n+1/2)} e^{-n}(1 + f(n))$, $\forall n$ entier, où $\lim_{n\to\infty} f(n) = 0$, l'on trouve :

$$\begin{aligned} -\log p_{\mathrm{N}}(\mathscr{C}_j) = & \sum_{i=1}^{\lambda}\left(n_i(j, \mathrm{N}) \log \left(\frac{n_i}{\mathrm{N}}\right)_j - n_i(j, \mathrm{N}) \log p_i\right) \\ & - \frac{1}{2}\left[\log \mathrm{N} - \sum_{i=1}^{\lambda} \log n_i(j, \mathrm{N}) + (1 - \lambda) \log 2\pi\right] - \log (1 + f(\mathrm{N})) \end{aligned}$$

$$+\sum_{i=1}^{\lambda}\log(1+f(n_i(j,\mathrm{N}))=\sum_{i=1}^{\lambda}n_i(j,\mathrm{N})\log\left(\frac{n_i}{\mathrm{N}}\right)_j$$
$$-\sum_{i=1}^{\lambda}n_i(j,\mathrm{N})\log p_i+\mathrm{T}(\lambda,\mathrm{N},j) \tag{3}$$

où $\left(\frac{n_i}{\mathrm{N}}\right)_j = \frac{n_i(j,\mathrm{N})}{\mathrm{N}}$ et $\mathrm{T}(\lambda, \mathrm{N}, j)$ désigne la somme des cinq derniers termes, qui dépend des paramètres λ, N, j. Plus loin, nous reviendrons brièvement sur la structure de (3). Mais pour l'instant formons tout de suite le rapport

$$\frac{-\log p_{\mathrm{N}}(\mathscr{C}_j)}{\mathrm{N}}=\sum_{i=1}^{\lambda}\left(\left(\frac{n_i}{\mathrm{N}}\right)_j\log\left(\frac{n_i}{\mathrm{N}}\right)_j-\left(\frac{n_i}{\mathrm{N}}\right)_j\log p_i\right)+\frac{\mathrm{T}(\lambda,\mathrm{N},j)}{\mathrm{N}} \tag{4}$$

Nous voulons examiner le comportement de ce rapport lorsque $\mathrm{N}\to\infty$. Pour un tel examen le métaespace $(\mathscr{U}^{\mathrm{N}}, \tau_t^{\mathrm{N}}, p_{\mathrm{N}}(\tau_t^{\mathrm{N}}))$ de puissance N *fixée* cesse évidemment d'être suffisant. Nous allons donc considérer un ensemble infini de tels métaespaces, un pour chaque valeur de N. La mesure $p_{\mathrm{N}}(\mathscr{C}_j)$ sera considérée, pour chaque N, dans le métaespace correspondant à *cet* N (2).

Remarquons maintenant qu'il est possible de faire en (4) tendre N vers l'infini en *fixant* la complexion $\mathscr{C}_j = \{f_{1j}\dots f_{ij}\dots f_{1j}\}$. En effet, soit $\mathscr{S}_j(\mathrm{N}) = (\mathrm{N}_{1j} < \mathrm{N}_{2j} < \dots < \mathrm{N}_{\mathrm{K}j} < \dots)$ une suite croissante indéfinie d'entiers tels que l'on puisse avoir

$$\forall\mathrm{N}_{\mathrm{K}j},\left[(\exists n_{i\mathrm{K}},\forall i):\frac{n_{i\mathrm{K}}}{\mathrm{N}_{\mathrm{K}j}}=f_{ij}\in\mathscr{C}_j\quad,\quad\sum_i n_{i\mathrm{K}}=\mathrm{N}_{\mathrm{K}j}\right]$$

Si alors N s'accroît vers l'infini en ne prenant que des valeurs $\mathrm{N}_{\mathrm{K}j}\in\mathscr{S}_j(\mathrm{N})$, la statistique $\mathscr{C}_j = \{\dots f_{ij}\dots\}$ peut être maintenue invariante, ce que nous allons indiquer en écrivant lim $\mathrm{N}\to\infty/j$ fixée.

---

(2) Cette approche peut paraître lourde, comparée à celle pratiquée dans certains traitements de McMillan ou de Khinchin, où *tous* les événements $\forall\mathrm{N}, \sigma_x^{\mathrm{N}}$, qui dans le métaespace $\left(\mathscr{U}^{\mathrm{N}}, \tau_t^{\mathrm{N}}, p_{\mathrm{N}}(\tau_t^{\mathrm{N}})\right)$ de puissance N finie correspondante sont élémentaires — sont intégrés, à la fois, dans un métaespace unique de puissance infinie $\left(\mathscr{U}^{\infty}, \tau_t^{\infty}, p_{\infty}(\tau_t^{\infty})\right)$ en tant qu'événements *non élémentaires*, *via* le concept de cylindre. Mais en fait, dans cette première étape de notre développement, une telle intégration est inappropriée pour notre but, car nous voulons d'abord étudier la probabilité $p_{\mathrm{N}}(\mathscr{C}_j)$ de la manière la plus simple et la plus intuitive. En outre, une telle intégration masquerait la distinction — importante — entre le concept de phénomène aléatoire $(\mathscr{U}^{\mathrm{N}}, \mathscr{P}^{\mathrm{N}})$ dérivé d'un phénomène aléatoire fondamental $(\mathscr{P}, \mathscr{U})$ et le concept de « source » au sens de la Théorie de l'Information, qui englobe les sources à mémoire nulle et les sources Markov.

Opérons un tel passage à la limite sur (4). Le premier terme de (4) reste invariant, tandis que le terme $T(\lambda, N, j)/N$ tend vers zéro pour tout λ et tout j comme on peut vérifier en examinant (3). En effet, chacun des cinq derniers termes de (3) — rapportés à N — tend séparément vers zéro lorsque N tend à l'infini, quels que soient λ et j. L'on a donc :

$$\lim N \to \infty / j \text{ fixée} \left(\frac{-\log p_N(\mathscr{C}_j)}{N} \right) = \sum_{i=1}^{\lambda} f_{ij} \log f_{ij} - \sum_{i=1}^{\lambda} f_{ij} \log p_i$$

$$= \sum_{i=1}^{\lambda} f_{ij} \log \frac{f_{ij}}{p_i} = \Omega(j/\{p_i\}) \quad (5)$$

où $\Omega(j/\{p_i\})$ est un réel dépendant de la loi fondamentale de probabilité $\{p_i\}$ et de la complexion statistique considérée j. Nous dénommons la quantité définie par (5) *la fonctionnelle d'opacité de la complexion statistique j par rapport à la loi fondamentale de probabilité* $\{p_i\}$, *en abrégé, l'opacité de j relativement à* $\{p_i\}$ (3). Le symbole $\Omega(j/\{p_i\})$ introduit en (5) désigne cette dénomination.

Le premier terme de la fonctionnelle d'opacité $\sum_{i=1}^{\lambda} f_{ij} \log f_{ij}$, est indépendant de la loi de probabilité fondamentale $\{p_i\}$, étant complètement déterminé par la seule structure statistique j en elle-même, qu'il caractérise. Ce terme possède à la fois une nature et une structure formelle analogues à celles de la grandeur $\sum_i n_i \log n_i$ qui, en thermodynamique statistique, a été associée à la « complexion de l'état thermodynamique », et dont le produit avec la constante k de Boltzmann, changé de signe,

$$-k \sum_i n_i \log n_i = S,$$

y a été dénommé l'entropie S de l'état thermodynamique.

Pour souligner ces analogies remarquables, nous introduisons la notation $S_j = -\sum_i f_{ij} \log f_{ij}$ et nous dénommons cette quantité *l'entropie abstraite de la statistique j* d'une séquence $\sigma_x^N \in \mathscr{U}^N$. Toutefois il faut garder présent à l'esprit le fait que la quantité $S_j = \sum_i f_{ij} \log f_{ij}$ est purement formelle,

(3) La base des log peut être quelconque. Par convention nous choisissons e.

non interprétée, tandis que l'entropie de Boltzmann est douée d'un degré plus réduit de généralité ayant été intégrée d'emblée à une description particulière (en outre elle fait intervenir les populations absolues n_i).

Le deuxième terme de la fonctionnelle d'opacité est foncièrement dépendant, à la fois, et de la statistique j et de la loi fondamentale de probabilité $\{p_i\}$: ce terme mesure la probabilité déterminée par $\{p_i\}$ pour l'*une* — quelconque — des suites σ_x^N ayant la statistique j, car on a

$$\sum_i f_{ij} \log p_i = \sum_i \frac{n_i(j, N)}{N} \log p_i = \frac{1}{N} \log \prod_i p_i^{n_i(j,N)} = \frac{1}{N} \log p_N(\sigma_x^N)_j.$$

Nous dénommons ce deuxième terme *la modulation* $\mathcal{M}(\{p_i\}/j)$ *de la loi de probabilité* $\{p_i\}$ *par la statistique* j. Avec les notations introduites (5) peut se réécrire

$$\Omega(j/\{p_i\} = -S_j - \mathcal{M}(\{p_i\}/j) \tag{5'}$$

V. PROPRIÉTÉS DE LA FONCTIONNELLE D'OPACITÉ

La fonctionnelle d'opacité (4) que nous venons de construire possède des propriétés qui nous semblent permettre une importante synthèse entre les concepts d'entropie statistique abstraite $S_j = -\sum_i f_{ij} \log f_{ij}$ de j, d'entropie probabiliste ou « informationnelle » $H = -\sum_i p_i \log p_i$ de la loi de probabilité sous-jacente, et la loi des grands nombres. Nous allons exprimer ces propriétés et cette synthèse à l'aide d'une suite de quatre théorèmes.

V.1. La relation entre l'entropie statistique abstraite et l'entropie probabiliste de Shannon. Signifiance d'une statistique.

THÉORÈME 1. — Étant donné un phénomène aléatoire $(\mathcal{P}, \mathcal{U})$ caractérisé par la loi de probabilité fondamentale $\{p_i\}$, pour la complexion $\mathcal{C}_{jF}$ définie par la statistique $\forall i, f_{ijF} = p_i$ la fonctionnelle d'opacité correspondante

$$\Omega(jF/\{p_i\} = \sum_i f_{ij,F} \log f_{ij,F} - \sum_i f_{ij,F} \log p_i = -S_{jF} - \mathcal{M}(\{p_i\}/jF)$$
$$= \sum_i p_i \log p_i - \sum_i p_i \log p_i.$$

a la valeur 0 qui est son minimum absolu.

Preuve. — Dénotons $f_{ij} = x_i$. Selon la méthode de Lagrange la condition d'extrémum de $\sum_i (x_i \log x_i - x_i \log p_i)$ sous la contrainte $\sum_i x_i = 1$ est équivalente à la condition

$$\forall s, \frac{\partial}{\partial x_s}\left(\sum_i x_i \log x_i - \sum_i x_i \log p_i\right) - \alpha \frac{\partial}{\partial x_s}\left(\sum_i x_i\right) = 0$$

où α est une constante arbitraire, et la contrainte $\sum_i x_i = 1$ étant respectée.

Ceci donne successivement :

$$\forall s, \log x_s + 1 - \log p_s - \alpha = 0$$

donc

$$\frac{x_s}{p_s} = \mathrm{K}^{\alpha-1} \quad , \quad \mathrm{K} \neq 1 \text{ étant la base des log (quelconque)}$$

donc

$$\left(\sum_{s=1}^{\lambda} x_s = \mathrm{K}^{\alpha-1}\sum_{s=1}^{\lambda} p_s = 1\right) \Rightarrow \mathrm{K}^{\alpha-1} = 1$$

Comme $\mathrm{K} \neq 1$ ceci entraîne $\alpha - 1 = 0$. Donc la condition d'extrémalité de $\Omega(j/\{p_i\})$ est

$$\forall s, \frac{x_s}{p_s} = 1, \qquad \text{donc} \qquad f_{sj} = p_s = f_{sj,\mathrm{F}} \tag{6}$$

Si l'on introduit la condition (6) dans (5), l'on trouve :

$$\begin{aligned}\Omega(j\mathrm{F}/\{p_i\}) &= \sum_i f_{ij,\mathrm{F}} \log f_{ij,\mathrm{F}} - \sum_i f_{ij,\mathrm{F}} \log p_i \\ &= \sum_i p_i \log p_i - \sum_i p_i \log p_i = 0 \end{aligned} \tag{7}$$

La solution (6) est unique, donc le minimum trouvé est absolu. Ceci toutefois peut être aussi démontré indépendamment :

Considérons maintenant une complexion j quelconque, et tenons compte du fait que la fonction $- x \log x$ est une fonction concave, qui donc satisfait à l'inégalité de Jensen : pour tout ensemble de λ valeurs u_i subdivisant un intervalle Δx et pour toute famille de λ constantes positives a_i, $i = 1, 2 \ldots \lambda$ telles que $\sum_{i=1} a_i = 1$, l'on a $- \log\left(\sum_i a_i x_i\right) \leq - \sum_i a_i \log x_i$. En appliquant cette inégalité au cas de la fonctionnelle d'opacité, pour j quelconque, l'on trouve

$$\left[\Omega(j/\{p_i\}) = \sum_i (f_{ij}\log f_{ij} - f_{ij}\log p_i) = -\sum_i f_{ij}\log\frac{p_i}{f_{ij}}\right]$$
$$\geqslant \left[-\log\left(\sum_i f_{ij}\frac{p_i}{f_{ij}}\right) = -\log 1 = 0\right].$$

Donc la fonctionnelle d'opacité est non négative, i. e. 0 est son minimum absolu. Le théorème 1 est ainsi complètement démontré.

La complexion $\mathscr{C}_{j\mathrm{F}} = j\mathrm{F} = \{\ldots f_{ij,\mathrm{F}}\ldots\} = \{\ldots p_i\ldots\}$ d'extrémalité de l'opacité sera dénommée la *complexion fidèle*, l'indice F indiquant cette dénomination.

La relation entre les deux entropies, statistique et informationnelle. — Le théorème 1 montre que toute complexion $j = j\mathrm{F}$ qui diffère de la complexion fidèle (à $\{p_i\}$) est « non transparente » pour $\{p_i\}$, engendrant une modulation « distorsionnante » $\mathscr{M}(\{p_i\}/j)$ où en général $f_{ij} \neq p_i$, et une opacité non nulle $\Omega(j/\{p_i\} > 0$. Seulement lorsque la statistique j est « mise au point » sur la statistique fidèle $j\mathrm{F}$ elle devient transparente pour $\{p_i\}$ et l'entropie statistique devient *l'entropie statistique fidèle*

$$\mathrm{S}_j = -\sum_i f_{ij.\mathrm{F}}\log f_{ij,\mathrm{F}} = -\sum_i p_i\log p_i$$

qui engendre une modulation *identique* ou *fidèle*

$$\mathscr{M}(\{p_i\}/j) = \sum_i f_{ij,\mathrm{F}}\log p_i = \sum_i p_i\log p_i$$

Alors l'opacité $\Omega(j\mathrm{F}/\{p_i\} = -\mathrm{S}_{j\mathrm{F}} - \mathscr{M}(\{p_i\}/j\mathrm{F})$ s'annule en conséquence de l'identification — au signe près, et en valeur numérique seulement des facteurs de même indice i, mais *pas* en contenu sémantique — de S_j et de $\mathscr{M}(\{p_i\}/j)$, avec une même quantité, à savoir $-\sum_i p_i\log p_i$. Or, cette quantité n'est autre que l'entropie « informationnelle » (ou « probabiliste ») H que Shannon a définie en 1948, mais en elle-même, indépendamment de toute statistique correspondante. *La fonctionnelle d'opacité fait apparaître ce même concept d'entropie probabiliste* $\mathrm{H} = -\sum_i p_i\log p_i$, *en relation organique cette fois avec l'entropie statistique abstraite*

$S_j = - \sum_i f_{ij} \log f_{ij}$, tout en distinguant clairement ces deux concepts d'entropie l'un de l'autre et en les emplaçant l'un par rapport à l'autre : la fonctionnelle d'opacité agit comme un télémètre abstrait qui estime à l'aide de l'entropie d'une statistique j la distorsion que la loi de probabilité sous-jacente $\{ p_i \}$, nécessairement existante, subit dans l'image qu'en fournit cette statistique. Dans le fonctionnement de ce télémètre abstrait *l'entropie probabiliste a lerôle d'indicateur de l'annulation de cette distorsion*, rôle qui coïncide avec celui de sélecteur de la statistique fidèle.

La signifiance d'une statistique. — Ces considérations mettent en évidence un fait très important, que l'on peut d'ailleurs prévoir déjà à partir des expressions (3) et (4) : l'entropie statistique abstraite de toute complexion statistique j liée à un phénomène aléatoire $(\mathscr{P}, \mathscr{U})$ est une quantité foncièrement relative à une certaine loi de probabilité sous-jacente. Ce n'est évidemment pas dans la structure formelle de la fonction d'entropie abstraite S_j que cette relativité se manifeste, ni dans la valeur numérique de cette fonction. L'un et l'autre de ces deux aspects de S_j sont indépendants de la loi de probabilité fondamentale, puisque la fonction $S_j = - \sum_i f_{ij} \log f_{ij}$ peut être construite pour tout j à partir de la seule donnée de ce j. Mais c'est la « signifiance » de S_j, le contenu sémantique associable à cette fonction, qui simplement se dissout dès que l'on supprime la référence à la loi de probabilité sous-jacente. Cette affirmation mérite d'être soutenue par quelques brèves considérations.

Un phénomène aléatoire $(\mathscr{P}, \mathscr{U})$, $\mathscr{U} = \{ \varepsilon_i \}$, $i = 1, 2 \ldots \lambda$, étant donné, la loi de probabilité fondamentale $\{ p_i \}$ que $(\mathscr{P}, \mathscr{U})$ introduit par sa définition, exprime une relation entre la procédure fixe et réitérable $\mathscr{P}$ et l'univers $\mathscr{U} = \{ \varepsilon_i \}$ d'événements élémentaires correspondant à $\mathscr{P}$ que l'on considère. Cette relation peut être regardée comme un codage en termes des ε_i, de certaines propriétés structurales inhérentes à la procédure $\mathscr{P}$, considérée dans sa globalité (comme par exemple dans le cas d'un dé jeté sur une surface, la loi $p_i = \frac{1}{6} =$ const., avec $i = 1, 2 \ldots 6$, $\varepsilon_i \equiv$ (la face opposée à celle portant l'inscription i s'est posée sur la surface) est un codage en termes de ces ε_i de la symétrie cubique du dé et de la planéité supposée de la surface). Les ε_i ont donc le rôle d'éléments d'un langage. Comme tout langage, le langage basé sur les ε_i implique un morcellement de la description. Chacun des ε_i est manifestement insuffisant à lui seul pour refléter des propriétés structurales globales de $\mathscr{P}$. L'entier ensemble $\{ \varepsilon_i \} = \mathscr{U}$ constituant tout l'univers d'événements ε_i « introduit » par $\mathscr{P}$ est nécessaire pour pouvoir construire — selon le code $\{ p_i \}$ — une certaine transposition de caractères globaux de $\mathscr{P}$, en valeurs numériques de fréquences relatives f_{ij}

d'apparition d'événements ε_i. Pour chacune des fréquences relatives f_{ij}, le « modèle » — parcellaire et codé — est constitué par la valeur numérique de la probabilité $p_i \in \{ p_i \}$ ayant le même indice i. Mais en général dans une série finie de N essais ces modèles parcellaires codés ne sont pas reproduits sans écarts par les fréquences f_{ij} enregistrées. Et même lorsque $N \to \infty$, certains obstacles (qualifiés comme « bruits ») peuvent empêcher la reproduction des p_i par les f_{ij}. La statistique j formée n'est donc pas en général rigoureusement fidèle à $\{ p_i \}$. Son degré plus ou moins grand de fidélité est mesuré par l'opacité $\Omega(j/\{ p_i \})$ de j par rapport à $\{ p_i \}$ qui définit ainsi le degré de connexion entre $\{ p_i \}$ et j et aussi l'efficacité de j comme base de prévisions concernant les occurrences futures des ε_i. La « signifiance » d'une statistique j liée à un phénomène aléatoire $(\mathscr{P}, \mathscr{U})$, pour autant qu'il n'est pas vide de sens de dire qu'une telle signifiance existe, consiste précisément et exclusivement dans ce degré de fidélité, et celui-ci est — foncièrement — une propriété de j relative à $\{ p_i \}$. La validité de cette conclusion pourrait paraître limitée parce que la catégorie de statistiques envisagée pourrait paraître particulière. L'on peut en effet se donner directement un univers $\mathscr{U} = \{ \varepsilon_i \}$, $i = 1, 2 \ldots \lambda$ d'événements ε_i et enregistrer une statistique f_{ij} des ε_i mais sans spécifier explicitement une procédure $\mathscr{P}$, donc sans définir un phénomène aléatoire $(\mathscr{P}, \mathscr{U})$ correspondant. En esprit positiviste l'on pourrait en outre postuler qu'aucune procédure $\mathscr{P}$ correspondant à $\mathscr{U} = \{ \varepsilon_i \}$ n'existe, et l'on pourrait affirmer que la statistique j possède néanmoins une signifiance prévisionnelle concernant des occurrences futures des ε_i, qui apparaît alors comme intrinsèque à j. Mais en quoi une telle signifiance intrinsèque pourrait-elle consister ? Imaginons que la statistique j que l'on avait enregistrée s'avère par la suite instable, i. e. que sa reproductibilité lors de réenregistrements de statistiques des ε_i s'avère faible. L'on dira alors que la statistique j « n'est pas significative », qu'elle n'a représenté qu'une « fluctuation », qui n'est pas une base efficace de prévisions sur les occurrences futures des ε_i. L'on cherchera donc une autre statistique $j' \neq j$ suffisamment stable, i. e. suffisamment identifiable à une certaine loi de probabilité $\{ p_i \}$ telle que l'opacité $\Omega(j'/\{ p_i \})$ soit négligeable. Ainsi la relativité à une loi de probabilité, de la signifiance d'une statistique, resurgit irrépressiblement. Sur la base de ces remarques nous concluons que — en toute généralité — la notion d'une signifiance intrinsèque d'une statistique j est une illusion et un non-sens. Cette notion manifeste une hâte de la pensée qui efface la perceptibilité de certains traits essentiels de la conceptualisation probabiliste, comme dans l'expérience des anneaux de Newton la vitesse de rotation des anneaux efface la perceptibilité des couleurs en ne laissant devant la vue qu'un blanc.

Toutes ces remarques se transfèrent évidemment à la fonction d'entropie statistique abstraite S_j d'une statistique j et elles montrent que l'idée d'une signifiance intrinsèque de cette fonction est elle aussi illusoire. En outre il semble bien clair que cette conclusion ne peut pas changer lorsqu'on considère, au lieu d'une statistique abstraite, une statistique d'événements ε_i réels et l'entropie statistique physique correspondante. Pourtant en physique l'on emploie depuis 100 ans les concepts d'entropies physiques, d'une manière absolutisée, sans aucune référence à une loi de probabilité spécifiée. Un tel usage ne peut être qu'une source de flottements et de faux problèmes, puisqu'il suggère une auto-suffisance sémantique qui n'est qu'un faux-semblant.

V.2. La probabilité d'une statistique, les deux entropies — statistique et probabiliste — et la loi des grands nombres

La Physique entretient encore une autre confusion grave concernant le concept d'entropie statistique, à savoir la croyance que l'entropie d'une statistique mesurerait la probabilité de réalisation de cette statistique (par l'intermédiaire du logarithme de cette probabilité). Or un bref examen des relations (2) $\to$ (5) suffit pour constater que sur le plan abstrait cette croyance est fausse, aussi bien pour tout N fini, qu'à la limite $N = \infty$.

Les relations (3) et (4) montrent que pour tout N fini

$$\forall j, \; - \log p_N(\mathscr{C}_j) \neq \sum_i f_{ij} \log f_{ij} \tag{8}$$

Si l'on considère alors la limite $N = \infty$, la relation (5) montre que

$$\forall j, \lim N \to \infty / j \text{ fixée} \left(\frac{- \log p_N(\mathscr{C}_j)}{N} \right) = \sum_i f_{ij} \log f_{ij} - \sum_i f_{ij} \log p_i \tag{8'}$$

ou le terme de modulation $\sum_i f_{ij} \log p_i$ n'est jamais nul. En outre, en (8′) la probabilité $p_N(\mathscr{C}_j)$ n'est plus individualisable, elle est intégrée organiquement dans la quantité $\lim N \to \infty / j \text{ fixée} \left(\frac{- \log p_N(\mathscr{C}_j)}{N} \right)$. L'on voit donc qu'il n'existe aucune relation simple entre l'entropie S_j d'une statistique j et sa probabilité $p_N(\mathscr{C}_j)$.

Ces différents faits formels ne font que manifester par des comportements mathématiques une raison conceptuelle profonde qui oblige à exclure *par principe* l'idée selon laquelle la fonction d'entropie d'une statistique pourrait à elle seule caractériser la probabilité de cette statistique. Cette raison est reliée étroitement à nos remarques précédentes sur la signifiance d'une statistique et de son entropie.

La fonction $S_j = - \sum_i f_{ij} \log f_{ij}$ d'entropie d'une statistique abstraite, à elle seule, simplement n'est pas reliée à la loi de probabilité fondamentale. La statistique j étant donnée, S_j reste la même quelle que soit la loi de probabilité $\{p_i\}$ sous-jacente. Donc S_j caractérise j *en elle-même*, sans référence à la loi de probabilité qui a agi. Mais une statistique en elle-même, de même qu'elle ne possède aucune signifiance, ne possède non plus aucune probabilité définie, ni de se réaliser, ni de ne pas se réaliser. Le concept d'une probabilité d'une statistique j, en absence de la spécification d'une loi de probabilité sous-jacente, est un faux absolu et un non-sens. La probabilité $p_N(\mathscr{C}_j)$ d'une statistique j — foncièrement — est une grandeur de relation entre la statistique j et la loi de probabilité $\{p_i\}$ sous l'action de laquelle j émerge : la statistique uniforme, par exemple, a selon la loi des grands nombres, la probabilité *maximale*, face à la loi $\{p_i\}$ uniforme, *cependant* que face à la loi $\{p_i\}$ δ, i. e. $p_k = 1$, $p_{i \neq K} = 0$, $\forall i \neq k$, cette *même* statistique uniforme a la probabilité *minimale*. Cet exemple est crucial : l'entropie s_j de j ne caractérise *aucunement* la probabilité $p_N(\mathscr{C}_j)$.

Mais alors où se trouve la source de cette croyance si courante en physique, selon laquelle l'entropie d'une statistique mesurerait le logarithme de la probabilité de cette statistique ? Notre hypothèse est que la source de cette croyance consiste dans la perception confuse de certaines caractéristiques du cas particulier d'une loi fondamentale de probabilité à distribution uniforme, $p_i = \dfrac{1}{\lambda} = \text{const}$, $\forall i$, $i = 1, 2 \dots \lambda$. Nous reviendrons sur cette hypothèse à la fin de ce chapitre pour l'étayer plus.

Mais d'abord nous essayerons de comprendre plus clairement la relation générale entre la probabilité d'une statistique j et les deux fonctions d'entropie, l'entropie statistique abstraite S_j de j et l'entropie informationnelle H de la loi de probabilité $\{p_i\}$ sous-jacente. Nous montrerons que pour $N = \infty$ cette relation ne s'élucide que sur le niveau d'organisation des *ensembles* de statistiques j, où elle fait apparaître une connexion profonde entre la loi des grands nombres et la fonctionnelle d'opacité.

Commençons par considérer le cas des N finis. Pour tout N fini la relation entre $p_N(\mathscr{C}_j)$, S_j et H est établie par (3) et (4). Cette relation n'a rien de simple car le terme $T(j, N)$ est non nul, mais elle ne soulève non plus aucun problème conceptuel. L'on peut constater que $p_N(\mathscr{C}_j) \neq 0$ pour tout j qui est compatible avec N. Ceci entraîne, *via* la loi des probabilités totales, que la probabilité $p_N(\mathscr{C}_{jF})$ de la statistique fidèle est moindre que 1 (en effet l'on voit aussi sur (3) que pour $j = jF$, bien que le premier terme « d'opacité » soit nul, le second terme $T(j, N)$ est non nul, donc $-\log p_N(\mathscr{C}_{jF}) \neq 0$ donc $p_N(\mathscr{C}_{jF}) \neq 1$). Tant que N est fini ce résultat confirme l'intuition. Mais lorsqu'on s'avance maintenant vers la limite $N = \infty$, il apparaît un problème qui conduit à une illustration intéressante de la dynamique de hiérarchisation des types logiques.

De (3) l'on obtient que

$$\forall j, \lim N \to \infty / j \text{ fixée } (-\log p_N(\mathscr{C}_j)) = \infty$$

donc (8″)

$$\forall j, \lim N \to \infty / j \text{ fixée } (p_N(\mathscr{C}_j)) = 0$$

Ainsi dans le métaespace de puissance $N = \infty$, $(\mathscr{U}^\infty, \tau_t^\infty, p_\infty(\tau_t^\infty))$, toute statistique j définie rigoureusement, individualisée parfaitement par rapport à toutes les autres statistiques $j' \neq j$, devient un événement de mesure de probabilité nulle. Et la statistique fidèle ne fait *pas* exception. Pourtant la loi des grands nombres affirme que la loi fondamentale de probabilité $\{ p_i \}$ obtient presque sûrement une définition fréquentielle à la limite $N = \infty$. L'on pourrait alors s'attendre sur la base de la loi des grands nombres, de trouver que pour la statistique fidèle $j = j\text{F}$ la probabilité $p_N(\mathscr{C}_j)$ de (3) et (4) tende vers 1 lorsque $N \to \infty$. Or (8″) montre que ceci n'est pas vrai.

Dès que cette conclusion s'est imposée ainsi mathématiquement, elle retient l'attention, et aussitôt sa signification se fait jour :

Au fur et à mesure que N s'accroît vers l'infini, l'exigence que toutes les populations relatives $n_i(j, N)/N = f_{ij}$, $\forall i$, restent définies rigoureusement, sans aucun écart (pour tout N compatible avec la statistique considérée $j = \{ f_{ij} \}$, $i = 1, 2 \dots \lambda$), acquiert un caractère de plus en plus extrême qui progressivement supprime la réalisabilité de cette exigence, quel que soit j, fidèle à $\{ p_i \}$ ou non. Ceci est un fait, et il appelle sa propre expression. Le calcul des probabilités fait face à cet appel en engendrant la propriété (8″) selon laquelle $\lim N \to \infty / j \text{ fixée } \left(\frac{-\log p_N(\mathscr{C}_j)}{N} \right) = 0, \forall j$. Mais en conséquence de cette propriété, la mesure de probabilité $p_N(\mathscr{C}_j)$ de (3) et (4) cesse à la limite $N = \infty$ de pouvoir différencier l'un de l'autre les degrés de réalisabilité des différentes statistiques j. Ceci entraîne que la singularisation de la statistique fidèle affirmée par la loi des grands nombres — qui elle aussi est un fait et réclame une expression — n'est pas descriptible sur le même niveau d'organisation logique des ε_i sur lequel s'emplace la définition des statistiques j définies rigoureusement, individuellement. Pour accéder à la descriptibilité probabiliste de la spécificité de la statistique fidèle, affirmée par la loi des grands nombres, l'on se trouve ainsi contraint, lorsqu'on passe à la limite $N = \infty$, de passer en même temps sur un autre niveau logique, supraordiné à celui où l'on a défini les statistiques j individuelles. Il faut maintenant considérer des ensembles de statistiques j différentes mais « proches » l'une de l'autre, et se limiter à rechercher la spécificité de l'entier ensemble de j seulement, qui forment le voisinage de la statistique fidèle. Ce relâchement de l'exigence d'une caractérisation rigoureusement individuelle de la spécificité de la statistique fidèle, qui s'est ainsi imposé, n'est d'ailleurs qu'une manifestation du bien connu caractère non certain mais probable seulement de la loi des grands nombres. Nous sommes en présence d'une illustration frappante engendrée par le calcul des probabilités, de la dynamique générale d'hiérarchisation des types logiques.

Nous allons donc considérer maintenant des ensembles de statistiques j et nous allons montrer qu'il existe une relation étroite entre la loi des grands nombres et les propriétés de la fonctionnelle d'opacité.

Soit un métaespace $(\mathscr{U}^{N}, \tau_t^N, p_N(\tau_t^N))$. Dans ce métaespace, soit l'ensemble $\{ j : \Omega(j/\{ p_i \}) \leq \eta \} \in \tau_t^N$ de toutes les statistiques j telles que l'opacité correspondante $\Omega(j/\{ p_i \})$ est inférieure ou égale au réel η arbitrairement petit. Ce critère sépare un certain « voisinage » de la statistique fidèle, mesuré par η.

La mesure de probabilité $p_N(\tau_t^N)$ pose sur cet ensemble une certaine probabilité $p_N[\{ j : \Omega(j/\{ p_i \}) \leq \eta \}] = p_N(\Omega(j/\{ p_i \}) \leq \eta)$. Concernant *cette* probabilité — et non pas concernant $p_N(\mathscr{C}_{jF})$ — l'on peut affirmer le

THÉORÈME 2. — La loi des grands nombres est équivalente à la proposition selon laquelle, pour toute paire de deux réels η, δ arbitrairement petits il existe un N_0 tel que

$$\forall N \geqslant N_0, p_N(\Omega(j/\{ p_i \}) \leq \eta) \geqslant (1 - \delta) \tag{9}$$

Preuve. — Nous voulons montrer que la loi des grands nombres entraîne (9) et que (9) entraîne la loi des grands nombres. Il faut alors préalablement exprimer la loi des grands nombres en relation explicite avec l'espace de probabilité $(\mathscr{U}^{N}, \tau_t^N, p_N(\tau_t^N))$ où nous affirmons le théorème 2 :

Soient deux réels arbitrairement petits η et δ et soit i un indice d'événement $\varepsilon_i \in \mathscr{U}$. Considérons l'ensemble $\left\{ \sigma_x^N : \left| \left(\frac{n_i}{N}\right)_x - p_i \right| \leqslant \eta \right\}$ de toutes les suites $\sigma_x^N \in \mathscr{U}^N$ telles que la fréquence relative d'apparition en σ_x de ε_i, $\left(\frac{n_i}{N}\right)_x$, s'écarte en valeur absolue de p_i, de moins de η ou de η tout au plus. Cet ensemble est un événement de la tribu totale τ_t^N, donc $p_N(\tau_t^N)$ pose sur cet ensemble une certaine probabilité

$$p_N\left(\left\{ \sigma_x^N : \left| \left(\frac{n_i}{N}\right)_x - p_i \right| \leq \eta \right\} \right) = p_N\left(\left| \frac{n_i}{N} - p_i \right| \leqslant \eta \right)$$

(ou la seconde écriture constitue une renotation simplifiée de la première). La loi des grands nombres exprimée relativement au métaespace $(\mathscr{U}^{N}, \tau_t^N, p_N(\tau_t^N))$ admet la forme :

Pour toute paire de réels η, δ arbitrairement petits, il existe un N_0 tel que

$$\forall i, p_N\left(\left| \frac{n_i}{N} - p_i \right| \leq \eta \right) \geqslant (1 - \delta) \tag{10}$$

Montrons maintenant que (9) entraîne (10) et (10) entraîne (9). Nous commençons par montrer que (10) entraîne (9).

La fonction log x est continue. Donc, si l'on choisit de fixer le réel η, l'on a

$$\forall i,\ [\exists \eta_i : (|\ x - p_i\ | \leq \eta_i) \Rightarrow |\ \log x - \log p_i\ | \leq \eta] \tag{11}$$

Appliquons la loi des grands nombres (10), avec η_i de (11) et un δ' quelconque. Elle permet d'écrire

$$\eta_i,\ \forall \delta' \text{ fixé},\ \forall i,\ \left[\exists N_{0i} : p_N\left(\left|\frac{n_i}{N} - p_i\right| \leq \eta_i\right) \geqslant (1 - \delta'),\ \forall N \geqslant N_{0i}\right] \tag{12}$$

Soit $N_0 = \max\limits_{\forall i} \{ N_{0i} \}$. De (11) et (12) l'on tire

$$\forall i,\ \forall N \geqslant N_0,\ p_N\left(\left|\frac{n_i}{N} - p_i\right| \leqslant \eta_i\right) \leqslant p_N\left(\left|\log\frac{n_i}{N} - \log p_i\right| \leq \eta\right) \tag{13}$$

De (12) et (13) il s'ensuit que

$$\forall \eta,\ \delta' \text{ donnés},\ \left[\exists N_0 : p_N\left(\left|\log\frac{n_i}{N} - \log p_i\right| \leqslant \eta\right) \geqslant (1 - \delta'),\ \forall N \geqslant N_0\right] \tag{14}$$

Donc

$$\forall \eta,\ \delta' \text{ donnés},\ \left[\exists N_0 : p_N\left(\left|\log\frac{n_i}{N} - \log p_i\right| > \eta\right) \leqslant \delta',\ \forall N \geqslant N_0\right] \tag{14'}$$

Formons maintenant la quantité $\sum\limits_i \frac{n_i}{N}\left|\log\frac{n_i}{N} - \log p_i\right|$ et considérons l'événement défini sur $(\mathscr{U}^N, \tau_t^N, p_N(\tau_t^N))$ par l'inégalité

$$\sum_i \frac{n_i}{N}\left|\log\frac{n_i}{N} - \log p_i\right| > \left(\sum_i \frac{n_i}{N}\eta = \eta\right)$$

i. e. l'événement

$$\left\{\sigma_x^N : \sum_i \left(\frac{n_i}{N}\right)_x \left|\log\left(\frac{n_i}{N}\right)_x - \log p_i\right| > \eta\right\} \in \tau_t^N.$$

Cet événement possède dans $(\mathscr{U}^N, \tau_t^N, p_N(\tau_t^N))$ une certaine probabilité

$$p_N\left(\left(\sum_i \frac{n_i}{N}\left|\log\frac{n_i}{N} - \log p_i\right|\right) > \eta\right)$$

(notation simplifiée) que nous allons estimer. A cette fin, remarquons que l'on a l'implication évidente

$$\left[\sum_i \frac{n_i}{N}\left|\log\frac{n_i}{N} - \log p_i\right| > \left(\sum_i \frac{n_i}{N}\eta\right)\right] \Rightarrow \left[\exists i : \left|\log\frac{n_i}{N} - \log p_i\right| > \eta\right] \tag{15}$$

(tandis que l'implication inverse n'est pas vraie). En termes de probabilités ceci se traduit par l'inégalité

$$p_N\left(\sum_i \frac{n_i}{N}\left|\log\frac{n_i}{N}-\log p_i\right|>\eta\right)>p_N\left(\exists i:\left|\log\frac{n_i}{N}-\log p_i\right|>\eta\right) \quad (16)$$

Estimons maintenant la probabilité $p_N\left(\exists i:\left|\log\frac{n_i}{N}-\log p_i\right|>\eta\right)$. L'on a

$$\begin{aligned} p_N\left(\exists i:\left|\log\frac{n_i}{N}-\log p_i\right|>\eta\right) &= p_N\left[\left(\left|\log\frac{n_1}{N}-\log p_1\right|>\eta\right)\right. \\ &\quad \cup\left(\left|\log\frac{n_2}{N}-\log p_2\right|>\eta\right)\cup\ldots\cup\left.\left(\left|\log\frac{n_\lambda}{N}-\log p_\lambda\right|>\eta\right)\right] \\ &= p_N\left[\bigcup_{i=1}^{\lambda}\left(\left|\log\frac{n_i}{N}-\log p_i\right|>\eta\right)\right] \end{aligned} \quad (17)$$

Mais, dans tout espace de probabilité, l'on a $p\left(\bigcup_i A_i\right)\le\sum_i p(A_i)$, ou A_i sont des événements de l'espace. Donc l'on doit admettre, tenant compte de (17), que

$$p_N\left(\exists i:\left|\log\frac{n_i}{N}-\log p_i\right|>\eta\right)\le\sum_i p_N\left(\left|\log\frac{p_i}{N}-\log p_i\right|>\eta\right) \quad (18)$$

De (18) et (14′) et tenant compte que $i = 1, 2 \ldots \lambda$, il s'ensuit que

$$\forall\eta,\ \delta' \text{ donnés, } N\geqslant N_0,\ \sum_{i=1}^{\lambda} p_N\left(\left|\log\frac{n_i}{N}-\log p_i\right|>\eta\right)\leqslant\lambda\delta' \quad (19)$$

De (16), (18) et (19) l'on conclut maintenant que

$$\forall\eta,\ \delta' \text{ donnés, } N\geqslant N_0,\ p_N\left(\sum_i\frac{n_i}{N}\left|\log\frac{n_i}{N}-\log p_i\right|>\eta\right)\leqslant\lambda\delta' \quad (20)$$

D'autre part l'on a $\left(\frac{n_i}{N}\text{ étant non négatif}\right)$

$$\left|\sum_i\left(\log\frac{n_i}{N}-\log p_i\right)\frac{n_i}{N}\right|\leqslant\sum_i\frac{n_i}{N}\left|\log\frac{n_i}{N}-\log p_i\right| \quad (21)$$

De (20) et (21) il découle

$$p_N\left(\left|\sum_i\frac{n_i}{N}\left(\log\frac{n_i}{N}-\log p_i\right)\right|>\eta\right)\le\left[p_N\left(\sum_i\frac{n_i}{N}\left|\log\frac{n_i}{N}-\log p_i\right|>\eta\right]\right.$$
$$\leqslant\lambda\delta',\ \forall N\geqslant N_0 \quad (22)$$

Mais dans l'inégalité qui figure dans le premier membre de (22), le premier membre est l'expression de la fonctionnelle d'opacité $\Omega(j/\{p_i\})$. L'on a donc trouvé que

$$\forall \eta,\ \delta' \text{ donnés, } [\exists N_0 : p_N(\Omega(j/\{p_i\}) > \eta) \leqslant \lambda\delta',\ \forall N \geqslant N_0] \tag{23}$$

Si l'on pose maintenant $\delta' = \frac{\delta}{\lambda}$ et l'on renverse le sens des inégalités de (23) l'on obtient

$$\forall \eta,\ \delta,\ \exists N_0 : \forall N \geqslant N_0,\ p_N(\Omega(j/\{p_i\}) \leqslant \eta) \geqslant (1 - \delta) \tag{24}$$

Ceci établit que la loi des grands nombres (10) entraîne (9).

Il reste à démontrer que (9) entraîne la loi (10) des grands nombres. Selon un résultat obtenu par Czizsar [6] l'on peut écrire :

$$\left(\sum_i a_i = 1,\ \sum_i b_i = 1,\ a_i \geqslant 0,\ b_i \geqslant 0\right)$$

$$\Rightarrow \left(\sum_i a_i \log \frac{a_i}{b_i}\right) \geqslant \frac{\left(\sum_i |a_i - b_i|\right)^2}{2} \tag{25}$$

En posant $a_i = \frac{n_i}{N}$ et $b_i = p_i$, l'on a donc

$$\left(\Omega(j/\{p_i\}) = \sum_i \frac{n_i}{N} \log \frac{\frac{n_i}{N}}{p_i}\right) \geqslant \frac{\left(\sum_i \left|\frac{n_i}{N} - p_i\right|\right)^2}{2} \geqslant \frac{\left|\frac{n_i}{N} - p_i\right|^2}{2} \tag{26}$$

En termes de probabilités, cette inégalité entraîne que

$$\forall \eta',\ p_N(\Omega(j/\{p_i\}) \leq \eta') \leq \left[p_N\left(\frac{\left|\frac{n_i}{N} - p_i\right|^2}{2} \leq \eta\right) = p_N\left(\left|\frac{n_i}{N} - p_i\right|\right) \leq \sqrt{2\eta'}\right] \tag{27}$$

En posant $\eta' = \frac{\eta^2}{2}$ l'on trouve que la loi (10) des grands nombres est une conséquence de (9). *q. e. d.* (4).

---

(4) Faisons ici une remarque importante : la démonstration du théorème 2 ne fait pas intervenir la signification probabiliste (4) de la quantité

$$\Omega(j/\{p_i\}) = \sum_i f_{ij} \log f_{ij} - \sum_i f_{ij} \log p_i.$$

Or cette quantité peut aussi être définie *directement*, en elle-même (comme l'a été l'entropie statistique ou l'entropie probabiliste). Cette quantité $\Omega(j/\{p_i\})$ sera alors *coupée* de l'hypothèse 1 « d'indépendance » des $\varepsilon_i \in \mathcal{U}$, mais le théorème 2 continuera d'être valide. En ce sens *la validité du théorème 2 ne requiert pas l'hypothèse d'indépendance 1*. Ceci pourrait s'avérer fertile pour l'extension des conséquences de ce théorème aux suites de signaux émis par une « source » à mémoire quelconque, au sens de la Théorie de l'Information.

La formulation que nous avons donnée plus haut au théorème 2 est en quelque sorte statique, en ce sens qu'elle concerne *directement* un métaespace de puissance N « assez grande ». Mais ce théorème admet une reformulation « dynamique » équivalente. Soit $\sigma_x^{N_0+\Delta N} \in \mathscr{U}^{N_0+\Delta N}$ une suite de $N = N_0 + \Delta N$ événements $\varepsilon_i \in \mathscr{U}$, $i = 1, 2 \ldots \lambda$, obtenue à partir d'une suite initiale $\sigma_{x_0}^{N_0} \in \mathscr{U}^{N_0}$ par ΔN réitérations additionnelles de $(\mathscr{P}, \mathscr{U})$. Au fur et à mesure que cette suite $\sigma_x^{N_0+\Delta N}$ se construit, la complexion $\mathscr{C}(\sigma_x^{N_0+\Delta N}) = \mathscr{C}_{j(N)}$ de $\sigma_x^{N_0+\Delta N}$ *évolue* en général avec N, à partir de la complexion de départ $\mathscr{C}_{j0} = \mathscr{C}(\sigma_{x_0}^{N_0})$. L'on peut donc, au lieu de fixer conceptuellement *une* certaine complexion d'indice caractéristique j constant, comme nous l'avons fait pour définir le concept d'opacité, changer maintenant d'optique et considérer la *complexion évolutive* $\mathscr{C}_{j(N)}$. Or,

THÉORÈME 2'. — La loi des grands nombres est équivalente à la proposition selon laquelle lorsque $(N = N_0 + \Delta N) \to \infty$ l'opacité $\Omega(j(N)/\{p_i\})$ converge en probabilité vers zéro, quelle que soit la complexion de départ $\mathscr{C}_{j0} = \mathscr{C}(\sigma_{x_0}^{N_0})$.

Preuve. — Évidente, sur la base de (5) et des démonstrations des théorèmes 1 et 2.

Enfin, le théorème 2, 2' a une conséquence qui syntaxiquement est évidente mais dont la signification justifie un énoncé indépendant :

THÉORÈME 3. — $\mathscr{C}_{j1} \neq \mathscr{C}_{j0} = \mathscr{C}(\sigma_{x_0}^{N_0})$ étant une seconde complexion statistique d'événements $\varepsilon_i \in \mathscr{U}$, choisie à l'avance arbitrairement, presque sûrement $\lim\limits_{\Delta N \to \infty} \mathscr{C}(\sigma_x^{N_0+\Delta N}) = \mathscr{C}_{j1}$ si et seulement si l'on réalise une procédure $\mathscr{P}'$ (par un choix adéquat des « conditions aux limites » par exemple) telle que pendant les ΔN essais additionnels la loi de probabilité fondamentale du phénomène aléatoire $(\mathscr{P}', \mathscr{U})$ soit $\{p_{1i}\} : \forall i,\ p_{1i} = f_{ij1} = \left(\frac{n_i}{N}\right)_{j1}$ vis-à-vis de qui $\mathscr{C}_{j1}$ est la complexion fidèle.

Preuve. — Immédiate, si l'on tient compte du théorème 2, 2', car l'on aura alors pour $\forall\eta$, δ arbitrairement petits $p_N(\Omega(\mathscr{C}_j(N_0 + \Delta N)/\{p_{i1}\}) \leq \eta) \geqslant (1 - \delta)$.

Dans les conditions de validité du théorème 3, au cours des ΔN essais additionnels, l'entropie statistique croît *ou décroît* en valeur, presque sûrement, selon que l'entropie de la complexion de départ $S_{j0} = -\sum_i f_{ij0} \log f_{ij0}$ est, respectivement, plus petite ou plus grande que l'entropie $H = -\sum_i p_{1i} \log p_{1i}$ de la loi de probabilité $\{p_{1i}\}$ sous-jacente aux ΔN essais additionnels.

Enfin, et ceci est important, la proposition incluse dans le théorème 2′ qui y est affirmée équivalente à la loi des grands nombres, peut être démontrée *indépendamment de la loi des grands nombres*, directement sur la base de l'expression analytique (4) de la probabilité (3) :

THÉORÈME 4. — Lorsque $N = N_0 + \Delta N$ tend à l'infini, l'opacité $\Omega(j(N)/\{p_i\})$ converge en probabilité vers zéro quelle que soit la complexion de départ $\mathscr{C}_{j0}$.

Preuve. — Réécrivons (4) sous forme condensée :

$$\frac{-\log p_N(\mathscr{C}_j)}{N} = \Omega(j/\{p_i\}) + \frac{T(\lambda, j, N)}{N} \tag{4}$$

où (voir (3))

$$T(\lambda, j, N) = \frac{1}{2}\left[-\log N + \sum_{i=1}^{\lambda} \log n_i(j, N) - (1-\lambda)\log 2\pi\right] - \log(1 + f(N)) + \sum_{i=1}^{\lambda} \log(1 + f(n_i(j, N))) \tag{29}$$

Par convention, si $n_i(j, N) = 0$ nous posons $\log n_i(j, N) = \log 0 = 0$ (le contexte admet cette convention comme naturelle). Examinons les termes de (29). L'on remarque :

$$\forall i, j, N, \qquad n_i(j, N) \geqslant 0, \qquad \text{donc} \qquad \sum_{i=1}^{\lambda} \log n_i(j, N) \geqslant 0. \tag{30}$$

En outre,

$$\exists B \text{ réel positif} : \forall (j, N), f(N) \leq B \tag{31}$$

En effet, la fonction $f(N)$ provient de la formule de Stirling $n! = \sqrt{2\pi}.n^{n+1/2} \times e^{-n}(1 + f(n))$, $\forall n$, où elle est définie comme une fonction positive qui converge vers zéro quand $n \to \infty$. Cette fonction est donc bornée. En (31) B est l'une de ses bornes. L'on peut donc écrire

$$\begin{aligned} \forall (j, N) \quad & \log(1 + f(N)) \leq \log(1 + B) \\ & -\log(1 + f(N)) \geqslant -\log(1 + B) \end{aligned} \tag{32}$$

Puisque $f(n)$ est une fonction positive l'on a aussi

$$\forall i, f(n_i) \geqslant 0 \Rightarrow \sum_{i=1}^{\lambda} \log(1 + f(n_i(j, N))) \geqslant \sum_{i=1}^{\lambda} \log(1) = 0 \tag{33}$$

De (30), (32), (33) l'on trouve

$$\forall (j, N) \sum_{i=1}^{\lambda} \frac{\log(n_i(j, N))}{2} - \log(1 + f(N)) + \sum_{i=1}^{\lambda} \log(1 + f(n_i(j, N))) \geqslant -\log(1 + B) \tag{34}$$

De (29) et (34) l'on tire

$$\forall(\lambda, j, \mathrm{N}),\ \mathrm{T}(\lambda, j, \mathrm{N}) \geqslant \frac{1}{2}[-\log \mathrm{N} - (1-\lambda)\log 2\pi - 2\log(1+\mathrm{B})] \quad (35)$$

Introduisons la notation $\mathrm{K} = \frac{1}{2}[-(1-\lambda)\log 2\pi - 2\log(1+\mathrm{B})]$. Alors (35) se réécrit

$$\forall(\lambda, j, \mathrm{N}),\ \mathrm{T}(\lambda, j, \mathrm{N}) \geqslant -\frac{\log \mathrm{N}}{2} + \mathrm{K}. \quad (35')$$

Donc

$$\forall(\lambda, j, \mathrm{N})\ \frac{\mathrm{T}(\lambda, j, \mathrm{N})}{\mathrm{N}} \geqslant -\frac{\log \mathrm{N}}{2\mathrm{N}} + \frac{\mathrm{K}}{\mathrm{N}} \quad (36)$$

Quand $\mathrm{N} \to \infty$, le deuxième membre de (36) tend vers 0. Donc

$$\forall\varepsilon, \exists \mathrm{N}_0(\varepsilon) : \left| \frac{-\log \mathrm{N}}{2\mathrm{N}} + \frac{\mathrm{K}}{\mathrm{N}} \right| \leq \frac{\varepsilon}{2},\ \forall \mathrm{N} \geqslant \mathrm{N}_0(\varepsilon)$$

$$-\frac{\log \mathrm{N}}{2\mathrm{N}} + \frac{\mathrm{K}}{\mathrm{N}} \geqslant -\frac{\varepsilon}{2},\ \forall \mathrm{N} \geqslant \mathrm{N}_0(\varepsilon) \quad (37)$$

où ε est un réel positif arbitrairement petit. De (36) et (37) l'on tire

$$\forall\lambda, \forall\varepsilon, \exists \mathrm{N}_0(\varepsilon) : \frac{\mathrm{T}(\lambda, j, \mathrm{N})}{\mathrm{N}} \geqslant -\frac{\varepsilon}{2},\ \forall(j(\mathrm{N})),\ \forall \mathrm{N} \geqslant \mathrm{N}_0 \quad (38)$$

Ce résultat intermédiaire est le point central de cette démonstration : en effet, nous avons défini le concept d'opacité pour *une* complexion statistique j, quelconque mais *fixée* :

$$\lim \mathrm{N} \to \infty / j \text{ fixée} \left(\frac{-\log p_{\mathrm{N}}(\mathscr{C}_j)}{\mathrm{N}} \right) = \Omega(j/\{p_i\}) + \lim \mathrm{N} \to \infty / j \text{ fixée } \frac{\mathrm{T}(\lambda, j, \mathrm{N})}{\mathrm{N}}$$

car $\Omega(j/\{p_i\})$ reste *invariant* lorsque $\mathrm{N} \to \infty$, *tant que j est fixée.* D'autre part, pour tout N donné, le cardinal de l'ensemble des complexions $j(\mathrm{N})$ réalisables — λ étant donné — est $\mathrm{M}(\lambda, \mathrm{N}) = \frac{(\mathrm{N}+\lambda-1)\,!}{(\lambda-1)\,!\,\mathrm{N}\,!}$ [*15*]. Lorsque N s'accroît, l'ensemble des statistiques $j(\mathrm{N})$ possibles *varie.* Dans ces conditions il n'était nullement évident *a priori* que pour tout ε il soit possible de spécifier un entier $\mathrm{N}_0(\varepsilon)$ assurant l'inégalité $\mathrm{T}(\lambda, j, \mathrm{N})/\mathrm{N} \geqslant -\frac{\varepsilon}{2}$, $\forall \mathrm{N} \geqslant \mathrm{N}_0(\varepsilon)$ pour tous les $j(\mathrm{N})$ *à la fois.* Une fois ce point central établi, la suite de la démonstration est sans difficulté :

Considérons l'ensemble $\{\sigma_x^{\mathrm{N}} : \Omega(\mathscr{C}(\sigma_x^{\mathrm{N}})/\{p_i\}) \geqslant \varepsilon\}$ de toutes les suites $\sigma_x^{\mathrm{N}} \in \mathscr{U}^{\mathrm{N}}$ telles que l'opacité de leur statistique est supérieure ou égale à ε. Cet ensemble appartient à τ_t^{N}, c'est-à-dire il est un événement en

$(\mathcal{U}^N, \tau_t^N, p_N(\tau_t^N))$. Estimons sa probabilité $p_N(\{\sigma_x^N : \Omega(\mathcal{C}(\sigma_x^N)/\{p_i\}) \geqslant \varepsilon\}) = p_N(\Omega(j(N)/\{p_i\}))$. Chaque statistique $j(N)$ individualisée correspond en $(\mathcal{U}^N, \tau_t^N, p_N(\tau_t^N))$ à un ensemble $\{\sigma_x^N : \mathcal{C}(\sigma_x^N) = j(N)\} = \{j(N)\} \in \tau_t^N$ (renotation simplifiée), qui est un événement et un événement disjoint de tout autre événement du même type :

$$\{j(N)\} \cap \{j'(N)\} = \phi, \forall(j(N) \neq j'(N))$$

Donc la loi des probabilités totales exige

$$p_N(\Omega(j(N)/\{p_i\}) \geqslant \varepsilon) = \sum_{j(N):\Omega(j(N)/\{p_i\}) \geqslant \varepsilon} p_N(\{j(N)\}) \tag{39}$$

où la sommation porte sur toutes les complexions $j(N)$ possibles telles que leur opacité est inférieure ou égale à ε. Mais en conséquence de (4) et (38) l'on a

$$\forall N \geqslant N_0, \forall(j(N) : \Omega(j(N)/\{p_i\}) \geqslant \varepsilon), \frac{-\log p_N(\mathcal{C}_{j(N)})}{N} \geqslant \varepsilon - \frac{\varepsilon}{2} \tag{40}$$

c'est-à-dire

$$\forall N \geqslant N_0, \forall(j(N) : \Omega(j(N)/\{p_i\}) \geqslant \varepsilon), p_N(\mathcal{C}_{j(N)}) \leqslant e^{-\frac{\varepsilon N}{2}} \tag{41}$$

Ceci vaut pour les complexions *individualisées* $j(N) : \Omega(j(N)/\{p_i\}) \geqslant \varepsilon$ sur lesquelles porte la somme de (39). Le nombre de telles complexions est inférieur ou égal au nombre $M(\lambda, N)$ de *toutes* les complexions $j(N)$ possibles pour le N considéré (λ étant fixé). Donc de (39) et (41), l'on tire, concernant *l'ensemble* des complexions contribuant au membre droit de (39),

$$\forall N \geqslant N_0(\varepsilon), p_N(\Omega(j(N)/\{p_i\}) \geqslant \varepsilon) \leqslant M(\lambda, N).e^{-\frac{\varepsilon N}{2}} \tag{42}$$

En outre, l'on a évidemment :

$$\begin{aligned} M(\lambda, N) = \frac{(N+\lambda-1)\,!}{(\lambda-1)\,!\,N\,!} &= \frac{1.2.\ldots(N-1)N(N+1)\ldots(N+\lambda-1)}{(\lambda-1)\,!\;1.2.\ldots(N-1)N} \\ &= \frac{(N+1)\ldots(N+\lambda-1)}{(\lambda-1)\,!} < (N+\lambda-1)^{\lambda} \end{aligned} \tag{43}$$

De (43) et (42) l'on déduit

$$\forall N \geqslant N_0(\varepsilon), p_N(\Omega(j(N)/\{p_i\}) \geqslant \varepsilon) < (N+\lambda-1)^{\lambda}.e^{-\frac{\varepsilon N}{2}} \tag{44}$$

De (44) l'on peut finalement conclure que

$$0 \leq \lim N \to \infty\, [p_N(\Omega(j(N)/\{p_i\}) \geqslant \varepsilon)] \leqslant \left[\lim N \to \infty\, (N+\lambda-1)^{\lambda}.\,^{-\frac{\varepsilon N}{2}} = 0\right] \tag{45}$$

q. e. d.

L'implication (9) $\Rightarrow$ (10) démontrée pour le théorème 2 permet maintenant sur la base du théorème 4, de considérer *la loi des grands nombres comme une conséquence de la convergence en probabilité, de l'opacité, vers zéro.* Cette possibilité montre à quel degré le concept d'opacité est fondamental et auto-suffisant.

Les propriétés établies par les théorèmes 2, 2′, 3 et 4 nous paraissent intéressantes. En effet, toute mesure de probabilité admet une définition fréquentielle au sens de la loi des grands nombres. Cette définibilité de principe n'assure pas à elle seule la mesurabilité *effective* des probabilités, à l'aide de fréquences relatives. Une telle mesurabilité requiert des conditions supplémentaires (convergence suffisamment rapide, uniforme, et d'autres conditions). Cette question de la mesurabilité opérationnelle, *effective*, d'une mesure abstraite de probabilité, à l'aide de fréquences relatives, est certainement d'une grande importance épistémologique [*7*], [*8*], [*9*], [*10*], car il s'agit là du critère même qui délimite l'ensemble de toutes les applications interprétatives vérifiables de la théorie abstraite des probabilités. Mais il n'en reste pas moins que la définibilité fréquentielle de principe, au sens de la loi des grands nombres, est une condition préalable pour la mesurabilité fréquentielle effective d'une mesure abstraite de probabilité. Elle possède donc, *a fortiori*, une importance épistémologique primordiale. Or, les théorèmes 2, 2′, 3 et 4 montrent qu'il existe des relations étroites entre la loi des grands nombres et les deux concepts d'entropie, l'entropie S_j d'une statistique abstraite $j(N)$ et l'entropie informationnelle H de la loi de probabilité sous-jacente à $j(N)$. En outre, la structure de ces relations se trouve entièrement explicitée, le long de l'entière trajectoire, de $N = 0$ à $N = \infty$ le long de laquelle la statistique $j(N)$ se constitue.

Lorsque N s'accroît, la statistique de toute suite σ_x^N tend à se soumettre à la loi des grands nombres *et* cette tendance est descriptible en termes d'une certaine dynamique abstraite que la structure analytique de la métaprobabilité $p_N(\mathscr{C}_j)$ — définie par (2), (3) et (4) — impose à la relation entre l'entropie statistique $S_{j(N)}$ et l'entropie probabiliste H. Cette dynamique est telle que l'entropie probabiliste invariante $H = -\sum_i p_i \log p_i$ — inobservable en elle-même comme une idée platonicienne — constitue pour l'entropie statistique $S_{j(N)} = -\sum_i f_{ij} \log f_{if}$ — observable et évolutive — *un modèle stable* ([5]) dont cette dernière construit progressivement, en fréquences relatives, une *image* de la structure formelle et de la valeur numérique (le contenu sémantique de H, *essentiellement* différent de celui de $S_{j(N)}$, n'est pas « reproduit », il échappe à l'expression morphique et numérique). Lorsque

([5]) En conséquence de l'identité de la procédure $\mathscr{P}$ réitérée.

cette image s'est constituée (presque), l'opacité $\Omega(j/\{p_i\})$ s'annule presque sûrement pour la statistique de très grande population totale N qui s'est alors réalisée, et désormais elle se maintient presque sûrement stationnaire autour de cette nullité. Comme *tout* processus de reproduction (ou « lecture ») progressive d'un modèle stable à l'aide d'entités *morcellantes* (« lettres » d'un « alphabet », puzzle, preneurs d' « empreintes » locales comme les molécules qui cognent des parois, etc.) ce processus d'évolution d'une entropie statistique $S_{j(N)}$ est lui aussi « fléché irréversiblement » : *vers son modèle* (6). Mais ce flèchement est dominé par deux relativités inamovibles, l'une dictée par la loi de probabilité sous-jacente, et l'autre dictée par la structure statistique de la complexion de départ, et ces deux relativités peuvent imposer à l'entropie de la statistique évolutive, une *décroissance* tout aussi bien qu'une croissance.

La loi des grands nombres décrit cette même dynamique d'une statistique évolutive $j(N)$, mais d'une manière tomisée, analysée, fréquence relative $f_{ij}(N)$ par fréquence relative. Or, par cette dissection tous les aspects globaux de l'évolution de $j(N)$ ont été filtrés, réduits, restant ainsi non perçus. *La fonctionnelle d'opacité forme pour la première fois une représentation précisément de ces aspects synthétiques de l'évolution d'une statistique* $j(N)$ *considérée dans sa globalité.* De même qu'en avion les contours et reliefs mouvants d'un nuage apparaissent clairement lorsqu'on sort de sa masse et d'au-dessus on le contemple dans la lumière du soleil, lorsqu'on s'impose le niveau d'examen qui permet une qualification globale d'une statistique évolutive $j(N)$, les concepts qui apparaissent avec une sorte d'évidence immédiate sont ceux de métaprobabilité $p_N(\mathscr{C}_{j(N)})$, d'entropie statistique $S_{j(N)}$ et d'entropie probabiliste $H(\{p_i\})$, et une connexion foncière, comme « organique », se révèle entre ces trois concepts. Les différences $|f_{ij(N)} - p_i|$ de la loi des grands nombres sont liées à ces concepts et à la fonctionnelle d'opacité qu'ils collaborent à définir, comme les molécules d'une structure sont liées à cette structure.

V.3. Le cas particulier d'une loi fondamentale de probabilité à distribution uniforme

Caractérisons maintenant brièvement le cas d'une loi de probabilité fondamentale $\{p_i\}$ à distribution uniforme, $p_i = \frac{1}{\lambda} = \text{const.}, \forall i, i = 1, 2 \ldots \lambda$.

L'expression (2) conduit immédiatement à la

(6) Ces résultats nous semblent confirmer fortement la conception de K. R. Popper [9] selon laquelle toute évolution d'une statistique manifeste une « propension » exprimée par la loi probabilité qui agit.

PROPOSITION 1. — Si $\{p_i\}$ est uniforme l'on a

$$\forall j, p_N(\mathscr{C}_j) = (N\,!/\Pi_i n_i(j, N))\left(\frac{1}{\lambda}\right)^N .$$

Donc en ce cas la permutabilité $W(j) = N\,!/\Pi_i n_i(j, N)$ de toute statistique i de population totale N finie caractérise à elle seule la probabilité $p_N(\mathscr{C}_j)$ de réalisation de j, à un facteur près $(1/\lambda)^N$ qui est indépendant de j.

La formule de Stirling permet d'affirmer que :

PROPOSITION 2. — Lorsque $N \to \infty$, la quantité $-\log W(j, N)/N$ tend vers $\sum_i f_{ij} \log f_{ij}$.

Le théorème 1 et la propriété bien connue de la fonction d'entropie d'atteindre son maximum pour une distribution uniforme, conduisent à la

PROPOSITION 3. — Si $\{p_i\}$ est uniforme, l'opacité est minimale pour la statistique d'entropie maximale.

Les théorèmes 2, 2' et 4 conduisent à la

PROPOSITION 4. — Si $\{p_i\}$ est uniforme, alors lorsque $N = N_0 + \Delta N$ tend à l'infini, l'entropie de la statistique évolutive $j(N)$ presque sûrement *croît* vers sa valeur maximale qui correspond à la statistique uniforme $f_{ij} = \frac{1}{\lambda} = \text{const.}, \forall i, i = 1, 2 \ldots \lambda$.

La relation (5) montre que

PROPOSITION 5. — Si $\{p_i\}$ est uniforme,

$$\lim N \to \infty/j \text{ fixée } (-\log p_N(\mathscr{C}_j)/N) = \sum_i f_{ij} \log f_{ij} - \log p_N$$

En ce cas donc l'entropie statistique abstraite $S_j = \sum_i f_{ij} \log f_{ij}$ mesure à elle seule la quantité $\lim N \to \infty/j$ fixée $(\log p_N(\mathscr{C}_j)/N)$ à une constante additive indépendante de j près.

La proposition 5 ne justifie évidemment *pas* l'affirmation que S_j mesurerait $\log p_N(\mathscr{C}_j)$ lorsque $\{p_i\}$ est uniforme, car en ce cas particulier — tout autant que dans le cas général de (8') — $\log p_N(\mathscr{C}_j)$ n'est pas individualisable de la quantité $\lim N \to \infty/j$ fixée $(-\log p_N(\mathscr{C}_j)/N)$.

L'ensemble des propositions 1, 2, 3, 4 semble étayer fortement l'hypothèse que nous avions avancée au début de ce chapitre : la croyance entretenue par la physique, selon laquelle l'entropie d'une statistique mesurerait le logarithme de la probabilité de cette statistique, semble refléter une perception implicite et partiellement erronée, exclusivement des caractéristiques du cas particulier d'une loi de probabilité $\{ p_i \}$ à distribution uniforme. Comme en ce cas l'expression générale (2)

$$p_N(\mathscr{C}_j) = (N\,!/\Pi_i n_i(j, N)\,!)\Pi_i p_i^{n_i(j,N)}$$

devient $p_N(\mathscr{C}_j) = (N\,!/\Pi_i n_i(j, N)\,!)(1/\lambda)^N$ où le facteur $(1/\lambda)^N$ ne dépend plus de j, l'on a naturellement tendance, lorsque l'attention se focalise sur ce cas seulement, à oublier le facteur $\Pi_i p_i^{n_i(j,N)}$. Il est alors compréhensible pourquoi l'on a baptisé le facteur de permutabilité $W(j, N) = N\,!/\Pi_i n_i(j, N)\,!$ à lui seul, « la *probabilité* thermodynamique de l'état » ([7]). Si ensuite implicitement on généralise cette même attitude au cas d'une distribution quelconque de $\{ p_i \}$, la relation entre j et $\{ p_i \}$ que (2) exprime, est coupée, et l'on se retrouve isolé dans les deux faux absolus d'une statistique j à signifiance intrinsèque et d'une probabilité intrinsèque $p_N(\mathscr{C}_j)$ d'une statistique j sans aucune référence à une loi de probabilité $\{ p_i \}$ sous-jacente. Dès lors il n'est nullement étonnant que l'on soit emporté sans défense dans des « paradoxes » et des faux problèmes. En effet, dans l'expression complète et générale (2) de $p_N(\mathscr{C}_j)$ le premier facteur de permutabilité n'est qu'un facteur de dénombrement, indépendant de $\{ p_i \}$. C'est le second facteur, $\Pi_i p_i^{n_i(j,N)} = p_N(\sigma_x^N)$, qui mesure la probabilité *définie par* $\{ p_i \}$ pour l'*une* (quelconque) des séquences σ_x^N ayant la statistique j, qui contient en germe toutes les relations entre j et $\{ p_i \}$. Il se déploie dans les expressions (3), (4) et (5) comme un fil d'Ariane marquant ces relations. Dans l'expression (5) de l'opacité il a engendré le second terme de modulation

$$\mathscr{M}(\{ p_i \}/j) = \sum_i f_{ij} \log p_i = \sum_i \frac{n_i(j, N)}{N} \log p_i$$
$$= \frac{1}{N} \log \Pi_i p_i^{n_i(j,N)} = \frac{1}{N} p_N(\sigma_x^N)_j$$

Et c'est par le terme de modulation de l'opacité, que se trouve assurée l'expression de la relation qui existe entre l'entropie S_j d'une statistique et la véritable probabilité (pas « thermodynamique ») de réalisation d'une statistique, déterminée par la structure j de cette statistique *et par* $\{ p_i \}$. Pour les N finis cette relation est exprimée par la formule (4). Pour $N \to \infty$, elle ne peut être recherchée que sur le niveau logique des ensembles de j, où s'emplace la loi des grands nombres.

Bien sûr ces considérations n'élucident pas entièrement la démarche des physiciens concernant les concepts d'entropie physique. Pourtant elles constituent une base de départ. Il sera important, dans un travail futur, d'expli-

---

([7]) L'ampleur de la confusion introduite par l'emploi du mot « probabilité » dans ce contexte peut être perçue à la lumière de l'exemple de la page 49 concernant la probabilité $p_N(\mathscr{C}_j)$ de la statistique j uniforme, face à une loi $\{ p_i \}$ uniforme d'une part, et face à une loi $\{ p_i \}$ à distribution δ d'autre part.

citer en détail la démarche des physiciens et d'établir les limites de validité des résultats qu'elle a produits.

V.4. La réalisation des opacités non nulles et leur utilité descriptive

Les commentaires du théorème 1 et ceux des théorèmes 2, 2′, 3 et 4 ont mis en évidence divers éclaircissements qu'apporte le concept d'opacité. Pourtant il existe un certain point de vue qui au premier abord semble réduire sérieusement l'intérêt de ce concept : l'opacité d'une statistique j n'est reliée à la probabilité $p_N(\mathscr{C}_j)$ d'une manière relativement simple, que pour $N = \infty$, selon la définition (5), *via* la quantité

$$\lim N \to \infty / j \text{ fixée} \left(\frac{- \log p_N(\mathscr{C}_j)}{N} \right).$$

Lorsqu'on explicite $p_N(\mathscr{C}_j)$ de (4) l'on trouve que pour $N \to \infty$, $p_N(\mathscr{C}_j) \to 0$, $\forall j$. Si alors, pour accéder à une descriptibilité comparative des différentes statistiques j l'on relâche la rigueur de la spécification des j en considérant des ensembles de j ayant des valeurs d'opacité voisines, l'on trouve — en accord avec la loi des grands nombres — que seules les statistiques ayant une opacité voisine de zéro se réalisent presque sûrement. Les autres j, presque sûrement, ne se réalisent pas. L'on peut alors se dire qu'après tout pour $N = \infty$ la distinction entre loi de probabilité et statistique devient un jeu inutile. Or c'est dans l'hypothèse $N = \infty$ que l'on se place pratiquement toujours. Du point de vue constitué par ces considérations, le concept d'opacité subit donc une éclipse. Mais si l'on continue l'examen la situation change aussitôt :

Une distribution $\{p_i\}$ étant donnée pour les $\varepsilon_i \in \mathscr{U}$, lors de mesures des fréquences relatives f_{ij} d'apparition des ε_i la loi des grands nombres ne se vérifie pas en général par rapport précisément à $\{p_i\}$. Car en général l'image de $\{p_i\}$ fournie par les f_{ij} est soumise à un « brouillage ». L'existence d'un tel brouillage peut s'exprimer en disant que la loi de probabilité effectivement agissante est une loi $\{p'_i\} \neq \{p_i\}$. Or, sur la base de la loi des grands nombres, ceci se traduit précisément par le fait que l'opacité *vis-à-vis de* $\{p_i\}$ de la statistique j enregistrée tend vers une valeur $K \neq 0$ lorsque $N \to \infty$. Mais vis-à-vis de $\{p'_i\}$ l'opacité de cette même statistique j doit être presque sûrement nulle (selon la loi des grands nombres et les théorèmes 2, 2′ et 4). Pour $N = \infty$ l'on a donc presque sûrement

$$\Omega(j/\{p_i\}) = \sum_i f_{ij} \log f_{ij} - \sum_i f_{ij} \log p_i = K$$

$$\Omega(j/\{p'_i\}) = \sum_i f_{ij} \log f_{ij} - \sum_i f_{ij} \log p'_i = 0$$

car presque sûrement $f_{ij} = p'$, $\forall i$. Donc presque sûrement il est justifié d'écrire

$$\Omega(j/\{p_i\}) = \sum_i p_i' \log \frac{p_i'}{p_i} = \mathrm{K}\ (^8) \tag{46}$$

L'on peut alors concevoir un « univers de bruit » $\mathscr{U}_b$ constitué par des événements élémentaires $\{\varepsilon_{l\to i}, l, i = 1, 2 \ldots \lambda\} = \mathscr{U}_b$ où $\varepsilon_{l\to i}$ a comme effet de substituer à l'événement $\varepsilon_l \in \mathscr{U}$ — qui se serait produit en son absence — l'événement $\varepsilon_i \in \mathscr{U}$. Si la loi de probabilité fondamentale correspondant à $\mathscr{U}_b = \{\varepsilon_{l\to i}\}$ est $\{p_{l\to i}\}$ alors on a évidemment $p_i' = p_i\left(1 - \sum_l p_{i\to l}\right) + \sum_l p_l p_{l\to i}$. En conjonction avec les traitements informationnels usuels des « voies avec bruit » — où l'on n'explicite pas une loi de probabilité sur l'univers de bruit — l'expression (46) fournit une base supplémentaire pour l'étude des « erreurs de transmission » (elle pourrait par exemple permettre des caractérisations utiles des distributions $\{p_{l\to i}\}$ à partir de l'enregistrement de j et de la donnée de $\{p_i\}$). Plus généralement, l'on peut établir à partir de la fonctionnelle d'opacité des expressions mathématiques utilisables comme des séparateurs de différentes lois de probabilité — supposées ou réelles — se composant de diverses manières.

Ces remarques nous semblent suffire pour soutenir l'affirmation que la fonctionnelle d'opacité possède probablement un grand intérêt pragmatique, précisément sur la base de la loi des grands nombres, et précisément en conséquence du fait qu'elle distingue entre probabilité et statistique, même pour $\mathrm{N} = \infty$.

VI. CONCLUSION

Nous avons construit un concept probabiliste nouveau, la fonctionnelle d'opacité $\Omega(j/\{p_i\})$ d'une statistique. A l'aide de quelques théorèmes nous avons établi les principales propriétés structurales et dynamiques de ce concept, pour tout N, de 0 à ∞. Ces propriétés font apparaître le fait que l'entropie statistique $\mathrm{S}_{j(\mathrm{N})} = -\sum_i f_{ij} \log f_{ij}$ d'une statistique abstraite $j(\mathrm{N})$, l'entropie informationnelle $\mathrm{H} = -\sum_i p_i \log p_i$ de la loi de probabilité

(8) L'on voit que la fonctionnelle d'opacité conduit ainsi tout naturellement et d'une manière justifiée conceptuellement, à la fonction de « gain d'information » introduite par Kullback [*11*].

$\{ p_i \}$ sous-jacente à j, et la loi des grands nombres, peuvent être considérées comme des aspects différents et emplacés sur des niveaux logiques distincts, du concept unique d'opacité de j par rapport à $\{ p_i \}$. Il nous semble certain que l'une des sources de la remarquable puissance descriptive qu'ont démontrée les deux concepts d'entropie, statistique et informationnelle, se trouve dans leurs relations avec la loi des grands nombres et avec les trois distributions de probabilité hiérarchisées $\{ p_i \}$, $p_N(\mathscr{C}_j)$, $p_N(j : \Omega(j/\{ p_i \}) \leq \eta)$, telles que ces relations s'établissent par rapport à la fonctionnelle d'opacité. La fonctionnelle d'opacité $\Omega(j/\{ p_i \})$ se révèle ainsi comme une entité descriptive centrale dans la conceptualisation probabiliste.

VII. PERSPECTIVES

En nous plaçant d'abord à l'intérieur de la théorie abstraite des probabilités, nous avons voulu doter les résultats d'une généralité maximale. Cette démarche peut être maintenant continuée en développant les résultats obtenus, sans encore quitter le niveau abstrait. Mais l'on peut également rechercher, dans les différentes descriptions probabilistes d'aspects de la réalité, des applications interprétatives des résultats abstraits déjà obtenus. Nous allons indiquer brièvement le contenu possible de ces deux voies.

VII.1. **Perspectives au niveau abstrait**

Les théorèmes 2′, 3, 4 suggèrent de nouvelles questions au niveau abstrait :

Que deviennent les lois d'évolution d'une entropie statistique abstraite $S_{j(N)}$ lorsque la statistique correspondante $j(N)$ se constitue sous l'action d'une loi de probabilité fondamentale $\{ p_i(\alpha, \beta \ldots) \}$ dépendant de certains paramètres ? En élaborant une réponse à cette question l'on pourrait préparer au niveau abstrait une théorie probabiliste des phénomènes qui mettent en jeu des « flux d'entropie » (au sens de Onsager [*12*], Prigogine [*5*]), en conséquence de l'existence de « gradients » de la loi fondamentale de probabilité.

En particulier, l'un des paramètres $\alpha, \beta \ldots$ pourrait s'identifier à N, qui est, sur le niveau abstrait, un équivalent désincarné — et essentiellement discret — du « temps » physique. Ou encore l'un des paramètres $\alpha, \beta \ldots$ pourrait s'identifier aux fréquences relatives f_{lj}, $\forall l$. En ces deux cas, la question précédente prend respectivement les formes : que deviennent les lois d'évolution d'une entropie statistique $S_{j(N)}$ si la statistique évolutive $j(N)$ se constitue sous l'action d'une loi fondamentale de probabilité $\{ p_i(N) \}$ variable elle-même avec N ? Que deviennent ces lois si la statistique $j(N)$ se constitue sous l'action d'une distribution fondamentale de probabilité

$\{ p_i(f_{lj}, \forall l) \}$ *qui subit une rétroaction* de la part de la statistique $j(\mathrm{N})$ déjà constituée ? Ces deux questions nous paraissent être liées à la préparation, au niveau abstrait, d'une biologie mathématisée, par exemple. Plus généralement, elles nous paraissent pouvoir préparer une théorie probabiliste des différentes évolutions organisatrices.

Quelles sont les probabilités *relatives* des différentes « trajectoires » $j(\mathrm{N})$, $\mathrm{N} \to \infty$? Sans doute que la famille des « trajectoires » $j(\mathrm{N})$, $\mathrm{N} \to \infty$ pour laquelle $\sum_{\mathrm{N}} \Omega(j(\mathrm{N})/\{ p_i \})$ est minimal, doit avoir la probabilité maximale. Ceci doit permettre une *théorie probabiliste de la stabilité structurale* (au sens de Thom [*13*] par exemple), une théorie probabiliste de l'observation (au sens de Vallée [*14*]) et la connexion avec l'approche de Feynman [*15*]. Plus généralement, *une théorie probabiliste de toute formulation variationnelle* devient envisageable à partir des théorèmes établis.

Enfin, l'ensemble de notre approche tient compte explicitement de la dépendance entre l'entropie d'une statistique, et l'univers $\mathscr{U} = \{ \varepsilon_i \}, i = 1, 2 \ldots \lambda$ d'événements élémentaires qu'une procédure $\mathscr{P}$ peut produire comme « effets ». Mais le cardinal λ de cet univers a été posé comme *fini*, donc *a fortiori* comme *borné*. Que deviennent les résultats obtenus si l'univers $\mathscr{U}$ est infini (discret *ou continu*) et si, en outre, il est *non borné*? Ces questions soulèvent des problèmes fondamentaux concernant d'une part l'entropie d'une distribution de probabilité fondamentale continue, et, d'autre part, concernant la signification et la *justification* de la condition de normation, imposée sans analyse à toute mesure d' « événement » (sous-ensemble mesurable de l'Univers). Ces questions ont été abordées déjà ailleurs [*16*], [*17*], mais elles méritent une analyse beaucoup plus achevée. Le problème si controversé de la « mort thermique de l'Univers », par exemple, que soulèvent les cosmologies, est commandé au niveau abstrait par l'examen de la justification de la condition de normation lorsque $\mathscr{U}$ est un infini continu et non borné. La même remarque vaut concernant les problèmes de « renormalisation » en théorie des particules élémentaires. L'entière approche probabiliste est à réexaminer dans ces cas.

Ces quelques remarques indiquent en quel sens les résultats présentés ici peuvent constituer une base de départ pour des développements que l'on doit élaborer d'abord à l'intérieur même de la théorie abstraite des probabilités, en vue d'applications interprétatives ultérieures.

VII.2. **Perspectives d'applications interprétatives à des descriptions probabilistes de la réalité**

Les relations mises en évidence par le nouveau concept d'opacité et par les théorèmes établis sont immanentes à la syntaxe de la conceptualisation probabiliste. Elles sont donc préalables logiquement à toutes les propriétés plus particulières introduites dans telle ou telle application interprétative de la théorie abstraite des probabilités, à la description d'un domaine de la réalité. Leur empreinte doit donc se retrouver sur toute telle application interprétative. Pour orienter la recherche de ces empreintes, il nous semble très utile de rappeler quelques aspects bien connus concernant l'élaboration des descriptions relativistes.

Un système formel, purement syntaxique, ne peut s'intégrer dans une théorie décrivant un domaine de la réalité, que *via* des postulats sémantiques affirmant que telle entité réelle est représentable par telle entité abstraite, qui alors transfère ses propriétés syntaxiques à cette entité réelle.

Dans le cas de la refonte relativiste, l'entité abstraite qui a joué le rôle central a été le concept de quadrivecteur covariant, défini à l'intérieur du calcul tensoriel : dans la mécanique newtonienne les coordonnées d'espace x, y, z d'un mobile et sa coordonnée de temps t, étaient déjà mises en relation de différentes façons (lois horaires, vitesse, accélération, etc.), mais elles n'étaient pas considérées comme des aspects d'une même entité. La même remarque s'applique à la masse inertielle et l'énergie, et à l'impulsion et l'énergie. L'entière restructuration einsteinienne a été basée sur le postulat *sémantique* nouveau selon lequel les coordonnées x, y, z, t d'un signal lumineux peuvent être considérées comme déterminant les 4 composantes d'une entité unique d'un certain type abstrait, à savoir un quadrivecteur covariant de norme invariante aux transformations de référentiel inertiel. Le concept abstrait de quadrivecteur *préexistait* en ce cas, mais le postulat sémantique mentionné comportait des relations *nouvelles* entre les grandeurs descriptives x, y, z, t, et celles-ci ont entraîné à leur tour des relations nouvelles entre la masse inerte et l'énergie, et entre l'impulsion et l'énergie. Ces dernières relations nouvelles ont conduit à l'ensemble de conséquences bien connu.

Considérons maintenant le concept d'opacité

$$\lim \mathrm{N} \to \infty / j \text{ fixée} \left(\frac{-\log p_{\mathrm{N}}(\mathscr{C}_j)}{\mathrm{N}} \right) = \Omega(j/\{p_i\})$$
$$= \sum_i f_{ij} \log f_{ij} - \sum_i f_{ij} \log p_i = -\mathrm{S}_j - \mathscr{M}(\{p_i\}/j)$$

A la différence de ce qui s'est passé dans le cas de la relativité, il s'agit ici d'une entité qui est nouvelle *en tant qu'entité abstraite* : dans la théorie abstraite des probabilités l'on admettait sans doute comme une évidence que la probabilité de réalisation d'une statistique est reliée à la loi de probabilité sous-jacente et à la structure de la statistique. Mais on n'avait jamais encore établi ces relations *explicitement*, et en faisant apparaître, par un choix convenable du niveau logique de description, l'entropie de la statistique et l'entropie probabiliste (en tant que grandeur caractérisant la valeur limite de l'opacité pour $j = j\mathrm{F}$). L'élément nouveau, dans ce cas, se situe donc *sur* le niveau purement abstrait. Il consiste précisément dans des relations *syntaxiques* mises en lumière entre la probabilité $p_N(\mathscr{C}_j)$ d'une complexion statistique j, l'entropie statistique S_j de j, et l'entropie probabiliste H de la loi de probabilité sous-jacente $\{ p_i \}$. Ces relations syntaxiques apparaissent maintenant comme une caractéristique nécessaire de toute complexion statistique j concernant des événements ε_i, $i = 1, 2, \ldots \lambda$ d'un univers $\mathscr{U}$-fini-quelconque. Elles doivent donc se transférer à toute statistique d'un ensemble $\{ \varepsilon_i \}$, $i = 1, 2 \ldots \lambda$ d'entités *réelles* qui — par postulat sémantique — est interprété comme l'univers d'événements élémentaires introduit par quelque phénomène aléatoire *réel*. En ce cas, le postulat sémantique mentionné, en lui-même, n'introduit *pas* d'élément nouveau; la source des conséquences nouvelles se trouve exclusivement dans le concept abstrait d'opacité. Toutefois, l'effet final doit être exactement du même type que dans le cas de la relativité restreinte : des entités décrivant un domaine de la réalité (des probabilités fondamentales attribuées à des événements réels, etc.) se trouveront soumises à des relations nouvelles, et ceci entraînera des conséquences.

Ces conséquences nous paraissent devoir être nombreuses : elles devraient conduire à des clarifications à l'intérieur de la Thermodynamique Statistique (9) et à l'intérieur de la Théorie de l'Information (10); elles devraient permettre une unification en profondeur des approches informationnelles avec les approches de la Thermodynamique Statistique; elles pourraient ouvrir la voie vers une Thermodynamique Statistique des processus irréversibles, et vers une théorie probabiliste des évolutions organisatrices du vivant; elles nous paraissent avoir des rapports étroits *avec les formulations variationnelles des lois* et en particulier avec les méthodes de Feynman. Ces différentes recherches, si elles étaient poursuivies, pourraient aboutir peut-être à une description probabiliste unitaire de toutes les évolutions réelles — désorganisatrices ou organisatrices, de la substance, des comportements,

---

(9) Lorsque l'entropie physique aura été clairement redéfinie comme une certaine interprétation du concept d'entropie statistique abstraite, la relativité de cette dernière, à une loi de probabilité fondamentale *et* à une entropie statistique de départ, conduira probablement, quand il en sera tenu compte explicitement, à l'identification de *limitations* de la validité du 2ᵉ principe de la thermodynamique. Corrélativement, certains « paradoxes » concernant le 2ᵉ principe pourraient apparaître comme des faux problèmes et se dissoudre. Par exemple, dans les exposés du « paradoxe » de Loschmidt l'on invoque le renversement des vitesses des molécules d'un gaz, comme une manipulation dont l'inefficacité pour reconduire à la statistique de départ devrait surprendre; or, la propriété abstraite exprimée par le théorème 3 suggère qu'une analyse mathématiquement correcte de la situation physique considérée devrait conduire à la conclusion que c'est la loi de probabilité fondamentale qu'il faudrait « renverser » en quelque sorte (en créant d'autres conditions aux limites), et *non pas les vitesses* (voir les notes 2 et 3).

(10) Les théorèmes de Birkhoff, McMillan, Khinchine, où il s'agit de « sources » et « entropie de la source », sans *référence explicite* à la loi de probabilité fondamentale des « lettres » émises, devraient s'éclaircir.

et même des connexions avec le réel, de l'esprit et des comportements. Enfin — et ceci est peut-être le plus important — il nous semble probable que le concept d'opacité et l'entière démarche qui nous a conduit à ce concept, permettront une représentation beaucoup plus analysée et claire de ce que l'on appelle « le temps ».

REMERCIEMENTS

Je voudrais tout d'abord remercier mon élève Nicolas Hadjisavvas qui tout au long de ce travail s'est révélé un interlocuteur remarquablement patient, perspicace et fin. Notamment il m'a aidée à dépasser certaines résistances dans les démonstrations.

Je suis reconnaissante au Professeur Raymond Payen d'avoir vérifié les démonstrations. Il a trouvé pour celles-ci une autre forme, beaucoup plus synthétique et élégante que celle qui est exposée ici, forme qui fera l'objet d'une publication à part.

Les Professeurs Costa de Beauregard et Robert Vallée ont accepté de lire le manuscrit et de me communiquer leurs opinions. Je les en remercie très vivement.

Je remercie vivement le Professeur H. Atlan de m'avoir communiqué des remarques très intéressantes qui m'ont conduite à élucider plus mes idées sur la « signifiance » d'une statistique, sur la relation entre statistiques et la loi de probabilité à la limite des grands nombres, et sur l'intérêt des opacités non nulles à la limite des grands nombres.

La vue intuitive d'où partent les constructions de ce travail a été probablement nourrie par la conception de K. R. Popper concernant la relation entre fréquences relatives et probabilités.

Je suis reconnaissante à mes élèves, le Docteur Daniel Evrard et François Thieffine, d'avoir lu le manuscrit et exprimé leurs critiques. Celles-ci ont conduit à une amélioration certaine de l'intelligibilité, notamment en ce qui concerne les notations.

NOTES

1. O. Costa de Beauregard écrit (*Sadi Carnot et l'essor de la thermodynamique*, C. N. R. S., 1976) ... « rendant très plausible l'idée que l'entropie des physiciens est réellement un défaut d'information et que, si elle peut être considérée comme une grandeur tout à fait objective lorsque les (très) grands nombres de l'échelle macroscopique sont en jeu, c'est de la même manière qu'une fréquence tend vers une probabilité ». En lisant ces lignes, nous traduisons ainsi : « l'entropie des physiciens » (qui sans doute est une certaine interprétation du concept abstrait d'entropie statistique $S_j = -\sum_i f_{ij} \log f_{ij}$) est un « défaut d'information » *face à l'entropie probabiliste* $H = -\sum_i p_i \log p_i$ qu'elle

arrive à reproduire lorsque $N \to \infty$, en devenant ainsi « objective », et en se soumettant au même mécanisme qui régit la loi des grands nombres en vertu de laquelle des fréquences relatives « déficientes » en « information » deviennent objectives lorsqu'elles égalent (presque) les probabilités (théorèmes 2, 2', 4). Une perception directe, intuitive, tellement exacte, de la relation entre S_j, H et la loi des grands nombres, réalisée en absence de tout support mathématique explicité, nous semble étonnante.

2. R. Lestienne (*Cahiers d'Histoire et de Philosophie des Sciences*, C. N. R. S., 1978, nº 9, p. 103) écrit :

... « l'entropie d'un système pourrait être une grandeur liée à l'interaction inévitable de tout système avec le monde extérieur, qui imposerait en quelque sorte son évolution ultérieure ». Cette formulation nous semble exprimer une idée voisine du théorème 3. En effet, dans notre approche le « monde extérieur » est représenté — en termes probabilistes — par la loi de probabilité fondamentale $\{ p_i \}$ *et* par la structure de la complexion statistique de départ $\mathscr{C}_{j0}$, que l'on peut concevoir — en principe — comme construite volontairement, avant de laisser se produire le processus d'évolution (isolée) de la statistique. Cette possibilité de principe est réalisable ou non selon le cas. Par exemple il ne semble pas impossible d'introduire dans un volume délimité par des parois thermostatées à une température très basse T, un gaz de $\mathcal{N}$ molécules préparées de manière à avoir des énergies cinétiques $1/2m_iv_i^2$ distribuées *uniformément* sur λ valeurs possibles $i = 1, 2 \ldots \lambda$, et de laisser ensuite ce gaz évoluer par interactions mutuelles et avec les parois. L'entropie probabiliste invariante correspondrait alors à la distribution de Maxwell-Boltzmann pour la température très basse (cette correspondance s'établit en fait par des considérations assez complexes), et elle serait plus petite que celle de la statistique initiale qui donc évoluerait en *décroissant*. Tous les processus « organisateurs » doivent être de ce type, i. e. correspondre à une paire $(\{ p_i \} \mathscr{C}_{j0})$ qui commande une décroissance de l'entropie statistique. Mais il semble impossible, au contraire, de manier les « conditions aux limites de l'Univers physique » (ce concept même, en outre, nous semble être un non-sens).

3. Le commentaire de la note 2 nous semble convenir également à plusieurs formulations très pertinentes de K. R. Popper, dont nous reproduisons une *in extenso* (*Nature*, t. **177**, 1956, p. 538) :

« It is widely believed that all irreversible mechanical processes involve an increase of entropy, and that « classical » (that is, non-statistical) mechanics, of continuous media as well as of particles, can describe physical processes only in so far as they are reversible in time. This means that a film taken of a classical process should be reversible, in the sense that, if put into a projector with the last picture first, it should again yield a possible classical process.

This is a myth, however, as a trivial counter-example will show. Suppose a film is taken of a large surface of water initially ar rest into which a stone is dropped. The reversed film will show contracting circular waves of increasing amplitude. Moreover, immediately behind the highest wave crest, a circular region of undisturbed water will close in towards the centre. This cannot be regarded as a possible classical process (it would demand a vast number of distant coherent generators of waves the coordination of which, to be explicable, would have to be shown, in the film, as originating from one centre. This, however, raises precisely the same difficulty again, if we try to reverse the amended film).

Thus irreversible classical process exist (on the other hand, in statistical mechanics all processes are, in principle, reversible, even if the reversion is highly improbable). Although the arrow of time is not implied by the fundamental equations, it nevertheless characterizes most solutions. For exemple, in applying Maxwell's theory, the initial conditions determine whether we choose to work with retarded or with advanced potentials, and the resulting situation is, in general, irreversible. »

La distinction entre équations et conditions aux limites est en effet fondamentale. Ce sont les conditions aux limites (et initiales), qui — en termes probabilistes — commandent le *sens* d'évolution, qui est exprimé par la solution mais pas par les équations. Pourtant il nous semble certain que ce type de débats ne pourra être véritablement tranché que lorsqu'on aura accompli jusqu'au bout un traitement probabiliste pur, et lorsqu'on aura — ensuite — traité les diverses applications interprétatives chacune séparément et à fond.

BIBLIOGRAPHIE

[*1*] M. Mugur-Schachter, *C. R. A. S.*, t. **288 A**, 2 mai 1979, p. 771.
[*2*] E. T. Jaynes, *Phys. Rev.*, t. **108**, 1957, p. 171; *Found. Phys.*, t. **3**, 1973, p. 477.
[*3*] M. Tribus, *Jour. Appl. Mechs.*, March 1961.
[*4*] A. Katz, *Principles of Statistical Mechanics*. W. H. Freeman, San Francisco, 1967.
[*5*] J. Prigogine, *Introduction à la Thermodynamique des Processus Irréversibles*, Dunod, Paris, 1968.
[*6*] Cziszar, *Stud. Scient. Math. Hung.*, t. **2**, 1967.
[*7*] R. von Mises, H. Geiringer, *The Mathematical Theory of Probability and Statistics*, N. Y. Acad. Press, 1964.
[*8*] R. Carnap, R. C. Jeffreys, eds., *Studies in Introductive Logic and Probability*, Univ. of Calif. Press, 1971.
[*9*] K. R. Popper, *La Logique de la Découverte Scientifique*, Paris, Payot, 1973.
[*10*] T. L. Fine, *Theories of Probability*, N. Y. and London, Acad. Press, 1973.
[*11*] Kullback, *Theory of Information and Statistics*, J. Wiley, N. Y., 1959.
[*12*] L. Onsager, *Phys. Rev.*, t. **37**, 1931, p. 405; t. **38**, 1931, p. 2265.
[*13*] R. Thom, *Stabilité Structurelle et Morphogenèse*, Benjamin, 1972.
[*14*] R. Vallée, *C. R. A. S.*, 1951, p. 1428; *C. R. A. S.*, t. **233**, 1951, p. 1350; Actes du VIII^{e} Congrès Int. de Cyb. Namur, 1973.
[*15*] R. P. Feynman, *Rev. Mod. Phys.*, t. **20**, 1948, p. 2.
[*16*] M. Mugur-Schachter, C. Padet, J. Padet, *C. R. A. S.*, t. **282**, A 487, mars 1976.
[*17*] N. Hadjisavvas, M. Mugur-Schachter, *Nuovo Cimento*.