Some Inequalities Satisfied by the Quantities of Information of Fisher and Shannon[1]

A. J. STAM

*Physics Laboratory of the Netherlands Defence Research Council,
The Hague, Netherlands*

A certain analogy is found to exist between a special case of Fisher's quantity of information I and the inverse of the "entropy power" of Shannon (1949, p. 60). This can be inferred from two facts: (1) Both quantities satisfy inequalities that bear a certain resemblance to each other. (2) There is an inequality connecting the two quantities. This last result constitutes a sharpening of the uncertainty relation of quantum mechanics for canonically conjugated variables. Two of these relations are used to give a direct proof of an inequality of Shannon (1949, p. 63, Theorem 15). Proofs are not elaborated fully. Details will be given in a doctoral thesis that is in preparation.

LIST OF SYMBOLS

$I(p)$	Fisher's quantity of information for the probability density $p(x_1, \cdots, x_n, \theta)$
$I_0(q)$	The special case of I arising for $n = 1$, $p(x,\theta) = q(x - \theta)$
$H(p)$	Shannon's quantity of information in natural units, referring to the probability density $p(x_1, \cdots, x_n)$. In Section 6, however, we denote it by $H(x_1, \cdots, x_n)$.
$N(p)$	Entropy power. $N(p) = \exp [2H(p)]/2\pi e$
$\sigma^2(p)$	Variance of the probability density $p(x)$
$\psi(x), \varphi(u)$	A pair of Fourier transforms [see (1.11)] with norm 1.
$\bar{z}$	Complex conjugate of z.

1. INTRODUCTION

Fisher's quantity of information (Fisher, 1925; Pitman, 1936) is defined by

$$I(p) = - \int p(x_1, \cdots, x_n, \theta) \frac{\delta^2}{\delta\theta^2} \ln p(x_1, \cdots, x_n\theta) \, dx_1 \cdots dx_n \quad (1.1)$$

---

[1] This investigation was suggested by Prof. J. L. van Soest, Director of Physical Research of the Netherlands Defence Research Council.

Here p is an n-dimensional probability density depending on the parameter θ. Under certain conditions (roughly speaking, if p is continuous and the boundaries of the x set where $p > 0$ do not depend on θ) we can write

$$I(p) = \int \frac{1}{p(x_1, \cdots, x_n, \theta)} \left(\frac{\delta p(x_1, \cdots, x_n, \theta)}{\delta \theta} \right)^2 dx_1 \cdots dx_n \quad (1.2)$$

It is well known that for independent variables I is additive: if

$$p(x_1, \cdots, x_n, \theta) = \prod_1^n p_j(x_j, \theta)$$

we have

$$I(p) = \sum_1^n I(p_j) \tag{1.3}$$

If p depends only on $x_i - \theta$ $(i = 1, \cdots, n)$ then θ drops from I. If, for $n = 1$, we have $p(x, \theta) = q(x - \theta)$, then

$$I(p) = \int \frac{1}{q(x)} q'^2(x) \, dx \tag{1.4}$$

where $'$ denotes differentiation with respect to x. In what follows we will use only the special case (1.4). We give it the notation $I_0(q)$. Any time we use this quantity it will be understood that q satisfies the conditions:

(i) $q > 0$ for $-\infty < x < \infty$

(ii) q' exists

(iii) The integral (1.4) exists, i.e., $q' \to 0$ rapidly enough for $x \to \pm \infty$.

These conditions also guarantee the equality of (1.1) and (1.2) and they are sufficient to allow the application of the theory of sufficient statistics as given by Fisher (1925) and Pitman (1936).

In this theory the following fact is proved: Let $T(x_1, \cdots, x_n)$ be any function of $x_1, \cdots, x_n$ and let $h(T, \theta)$ be the probability density for T as derived from the $x_1, \cdots, x_n$ distribution $p(x_1, \cdots, x_n, \theta)$. Then

$$I(h) \leqq I(p) \tag{1.5}$$

with equality if and only if

$$\frac{\delta \ln p(x_1, \cdots, x_n, \theta)}{\delta \theta} = F(T, \theta) \tag{1.6}$$

The quantity $I_0(q)$ is a measure for the "sharpness" of the distribution $q(x)$. If $q(x)$ is Gaussian, $I_0(q)$ is equal to $1/\sigma^2$ and if we "contract" the distribution q over a factor λ, $I_0(q)$ is multiplied by λ^2. The same facts hold for l/N, where N is the entropy power defined by $N = \exp(2H)/2\pi e$ (Shannon 1949, p. 69). This fact suggests some connection between these quantities, and part of this paper will be devoted to finding some more connections between them.

The quantity $I_0(q)$ can be written in a way that is known from the formalism of quantum mechanics. If we put $q(x) = |\psi(x)|^2$ with ψ a complex function with norm 1, we have

$$I_0(q) = 4 \int \left\{ \frac{d\,|\psi|}{dx} \right\}^2 dx \qquad (1.7)$$

But for any real u_0 we have also

$$I_0(q) = 4 \int \left\{ \frac{d}{dx} \left| \psi \exp(-2\pi i u_0 x) \right| \right\}^2 dx \qquad (1.8)$$

For any complex function f(z) of real x we have

$$\left(\frac{d\,|f|}{dx} \right)^2 = \frac{df}{dx}\frac{d\bar{f}}{dx} - |f|^2 \left(\frac{d}{dx}\arg f \right)^2$$

So

$$I_0(q) \leqq 4 \int \frac{d\psi \exp(-2\pi i u_0 x)}{dx} \frac{d\bar{\psi}\exp(2\pi i u_0 x)}{dx}\,dx \qquad (1.9)$$

with equality if and only if

$$\arg \psi = 2\pi i u_0 x + \text{const} \qquad (1.10)$$

If now $\psi(x)$ and $\varphi(u)$ are a pair of Fourier transforms:

$$\psi(x) = \int \varphi(u) \exp(2\pi i u x)\,du$$

$$\varphi(u) = \int \psi(x) \exp(-2\pi i u x)\,dx \qquad (1.11)$$

one finds that the right side of (1.9) is equal to $16\pi^2\sigma^2(|\varphi|^2)$ where

$$\sigma^2(|\varphi|^2) = \int (u - u_0)^2\,|\varphi(u)|^2\,du \qquad (1.12)$$

that is, $\sigma^2(|\varphi|^2)$ is the variance of the "momentum canonically conjugate to x." So

$$I_0(q) \leq 16\pi^2\sigma^2(|\varphi|^2) \qquad (1.13)$$

with the case of equality specified by (1.10).

2. STATEMENT OF RESULTS

A relation to be proved in the next sections is

$$I_0(p) \geq 1/\sigma^2(p) \qquad (2.1)$$

with equality if and only if p is Gaussian.[2] We have now to compare this fact with the maximizing of H by the Gaussian distribution expressed by

$$1/N(p) \geq 1/\sigma^2(p) \qquad (2.2)$$

The relation (2.1) can be sharpened by

$$I_0(p) \geq 1/N(p) \qquad (2.3)$$

Again, when p is Gaussian there is equality.

If we put $p = |\psi|^2$, as in Section 2, we can, making use of (1.13), set up a chain of inequalities in $|\psi|^2$ and $|\varphi|^2$:

$$1 \leq N(|\psi|^2)I_0(p) \leq \frac{\sigma^2(|\psi|^2)I_0(p)}{16\pi^2 N(|\psi|^2)\sigma^2(|\varphi|^2)} \leq 16\pi^2\sigma^2(|\psi|^2)\sigma^2(|\varphi|^2) \qquad (2.4)$$

The weakest of these inequalities is the usual formulation of the uncertainty principle. The sharpening

$$16\pi^2 N(|\psi|^2)\sigma^2(|\varphi|^2) \geq 1 \qquad (2.5)$$

is the most interesting for it suggests a further sharpening:

$$16\pi^2 N(|\psi|^2) \, N \, (|\varphi|^2) \overset{?}{\geq} 1 \qquad (2.6)$$

or[3]

$$H(|\psi|^2) + H(|\varphi|^2) \overset{?}{\geq} 1 - \ln 2 \qquad (2.7)$$

[2] Formula (2.1) is nothing but a special case of the Cramér-Rao inequality. However, it may be of interest in its connection with a sequence of sharpening formulations of the uncertainty principle (2.4).

[3] In fact the value $(1 - \ln 2)$ is attained for $\psi(x) = (2\pi\sigma^2)^{-1/4} [\exp(-x^2/2\sigma^2)]^{1/2}$.

However, the author has not been able to prove more than

$$H(|\psi|^2) + H(|\varphi|^2) \geqq 0 \tag{2.8}$$

Another inequality for I_0 is derived from the theory of sufficient statistics. Let $r(z)$ be the convolution of the probability densities $p_i(x_i)$, where $i = 1, \cdots, n$; that is, z is the sum of the independent random variables $x_1, \cdots, x_n$.

$$r(z) = \int p_1(x_1) \cdots p_{n-1}(x_{n-1}) p_n(z - x_1 - \cdots x_{n-1}) \, dx_1 \cdots dx_{n-1}$$

Then we have for any positive real numbers $\alpha_1, \cdots, \alpha_n$:

$$\left\{\sum_{i=1}^{n} \alpha_i\right\}^2 I_0(r) \leqq \sum_{i=1}^{n} \alpha_i^2 I_0(p_i) \tag{2.9}$$

with equality if and only if the p_i are Gaussian with variances proportional to the α_i:

$$\sigma^2(p_i) = c\alpha_i \tag{2.10}$$

There is a certain analogy, though less striking than in (2.1), to an inequality for N, viz., the inequality of Shannon[4] (1949, p. 63, Theorem 15)

$$N(r) \geqq \sum_{i=1}^{n} N(p_i) \tag{2.11}$$

We will make use of (2.9) to give a direct proof of (2.11). For this purpose we need another important relation. Let $p_v(x)$ be the convolution of a probability density $p(x)$ with a Gaussian probability density having variance v. Then

$$\frac{dH(p_v)}{dv} = \frac{1}{2} I_0(p_v) \qquad (v > 0) \tag{2.12}$$

The relations (2.3) and (2.12) were communicated to Prof. van Soest by Prof. N. G. de Bruijn who gave a variational proof of (2.3) and used (2.3) and (2.12) to give a direct proof of (2.11) for the case that all

[4] It was pointed out to the author by Prof. De Bruijn that (2.9) is equivalent to:

$$1/I_0(r) \geqq \sum_{i=1}^{n} 1/I_0(p_i)$$

which is a better analogy to (2.11).

but one of the p_i are Gaussian. We will reverse this order and derive (2.3) from (2.11).

All these results could be generalized to multidimensional distributions. Here this will be done only for (2.11) where we will not generalize the proof but show by induction that the inequality holds for multidimensional probability densities.

3. PROOF OF (2.1)

We follow the line of the proof of Theorem 226 of Hardy *et al.* (1952, p. 165). If h is a differentiable complex function of x on $(-\infty, \infty)$ with

$$\lim_{x \to \pm\infty} |\psi|^2 h = 0$$

we find by partial integration

$$\int |\psi|^2 h' \, dx = -2 \int h |\psi| |\psi|' \, dx$$

Applying the Schwarz inequality, we have

$$\left| \int |\psi|^2 h' \, dx \right|^2 \leq 4 \int |h|^2 |\psi|^2 \, dx \times \int |\psi|'^2 \, dx \qquad (3.1)$$

with equality if and only if

$$A |h|^2 |\psi|^2 = B |\psi|'^2$$

$$\arg h |\psi| |\psi|' = \text{const}$$

These conditions are equivalent to

$$|\psi|' = Ch |\psi|$$

or

$$|\psi(x)| = \alpha \exp (\beta h(x)) \qquad (3.2)$$

For $h = x - x_0$, (3.1) becomes something like Weyl's inequality (Weyl, 1949):

$$\left\{ \int |\psi|^2 \, dx \right\}^2 \leq 4 \int (x - x_0)^2 |\psi|^2 \, dx \times \int |\psi|'^2 \, dx$$

which by (1.7) goes over into (2.1).

4. PROOF OF (2.9)

In (1.5) we take $p(x_1, \cdots, x_n, \theta) = \prod_{i=1}^{n} q_i(x_i - \theta)$ and

$$T = \sum_{i=1}^{n} \alpha_i x_i.$$

Then one has

$$h(T, \theta) = r(T - \theta \sum_{i=1}^{n} \alpha_i)$$

where r is the convolution of the probability densities p_i defined by

$$p_i(x) = 1/\alpha_i q_i(x/\alpha_i) \qquad (i = 1, \cdots, n)$$

the relation (1.5) becomes now

$$\left\{ \sum_{i=1}^{n} \alpha_i \right\}^2 \int \frac{1}{r(z)} r'^2(z) \, dz \leqq \sum_{i=1}^{n} \int \frac{1}{q_i(x)} q_i'^2(x) \, dx$$

By substituting $q_i(x) = \alpha_i p_i(\alpha_i x)$ we arrive at (2.9). By (1.6) there is equality if and only if

$$\sum_{i=1}^{n} \frac{q_i'(x_i - \theta)}{q_i(x_i - \theta)} = F(\alpha_1 x_1 + \cdots + \alpha_n x_n, \theta)$$

From this condition can be derived the well known fact that the q_i have to be Gaussian with variances $\sigma^2(q_i) = C/\alpha_i$. As $\sigma^2(p_i) = \alpha_i^2 \sigma^2(q_i)$ the conditions in the p_i are

$$\sigma^2(p_i) = C\alpha_i$$

5. THE RELATION (2.12)

We have

$$p_v(z) = \int p(x)(2\pi v)^{-1/2} \exp\left(-(z - x)^2/2v\right) \, dx$$

By differentiation under the integral sign one sees that

$$\frac{dp_v(z)}{dv} = \frac{1}{2} \frac{d^2 p_v(z)}{dz^2} \qquad (v > 0) \quad (5.1)$$

from which (2.12) is derived by partial integration in

$$\frac{dH(p_v)}{dv} = -\int (1 + \ln p_v) \frac{dp_v(z)}{dv} \, dz$$

It is clear that p has to satisfy some conditions in order that the operations carried out above are allowed and p_v satisfies the conditions (i), (ii), (iii) of section 1. We shall not set up these conditions here but only remark that they can be much weaker than (i), (ii), (iii) of section 1. It is not required that p be differentiable or even continuous everywhere or that $p > 0$ on $(-\infty, \infty)$.

6. PROOF OF (2.11)

It is sufficient to prove (2.11) for $n = 2$. For higher n the inequality follows by repeated application of the case $n = 2$. Let p,q be probability densities and p_λ, q_λ the convolutions of p, q with Gaussian probability densities having variances $f(\lambda)$, $g(\lambda)$. The functions f and g depend on the parameter λ. So far we have only supposed that f and g are positive and have positive derivative for $\lambda > 0$ and that $f(0) = g(0) = 0$. Now consider the ratio

$$V(\lambda) = \frac{\exp[2H(p_\lambda)] + \exp[2H(q_\lambda)]}{\exp[2H(r_\lambda)]}$$

with

$$r_\lambda(z) = \int p_\lambda(x) q_\lambda(z - x) \, dx$$

Evidently

$$V(0) = \{N(p) + N(q)\}/N(r)$$

By (2.12) we have

$$\frac{dV(\lambda)}{d\lambda} \exp[2H(r_\lambda)] = f'(\lambda) I_0(p_\lambda) \exp[2H(p_\lambda)] + g'(\lambda) I_0(q_\lambda) \exp[2H(q_\lambda)]$$

$$- I_0(r_\lambda)\{f'(\lambda) + g'(\lambda)\} \{\exp[2H(p_\lambda)] + \exp[2H(q_\lambda)]\}$$

Applying (2.9) with

$$\alpha_1 = (f')^{1/2} \exp[H(p_\lambda)] \qquad \alpha_2 = (g')^{1/2} \exp[H(q_\lambda)]$$

we have

$$\frac{dV(\lambda)}{d\lambda} \exp[2H(r_\lambda)]/I_0(r_\lambda) \geqq \{(f')^{1/2} \exp[H(p_\lambda)] \tag{6.1}$$
$$+ (g')^{1/2} \exp[H(q_\lambda)]\}^2 - (f' + g')\{\exp[2H(p_\lambda)] + \exp[2H(q_\lambda)]\}$$

Now we choose the functions f and g so that

$$f'(\lambda) = \exp[2H(p_\lambda)] \qquad g'(\lambda) = \exp[2H(q_\lambda)] \qquad (6.2)$$

For this choice of f and g we have

$$dV(\lambda)/d\lambda \geqq 0$$

For any specified value of λ there is equality in (6.1) if and only if p_λ and q_λ are Gaussian. But then p and q have to be Gaussian and (6.1) is an equality for all λ. So $V(\lambda)$ is either strictly increasing or a constant. We still have to verify the conditions (2.10) for equality in (6.1):

$$\sigma^2(p_\lambda) = c(\lambda)\alpha_1 = c(\lambda)\exp[2H(p_\lambda)]$$

$$\sigma^2(q_\lambda) = c(\lambda)\alpha_2 = c(\lambda)\exp[2H(q_\lambda)]$$

They are satisfied by $c(\lambda) = 2\pi e$.

As $V(\lambda)$ is continuous from the right in $\lambda = 0$ we have

$$V(0) = \{\exp[2H(p)] + \exp[2H(q)]\}\exp[-2H(r)] \leqq \lim_{\lambda\to\infty}V(\lambda)$$

with equality if and only if p and q are Gaussian.

From (6.2) it is clear that

$$\lim_{\lambda\to\infty}f(\lambda) = \lim_{\lambda\to\infty}g(\lambda) = \infty$$

The fact that $\lim_{\lambda\to\infty}V(\lambda)$ exists and is equal to 1 can be proved easily, making use of the fact that p_λ, q_λ, r_λ "become more and more Gaussian."

7. EXTENSION OF (2.11) TO m-DIMENSIONAL PROBABILITY DENSITIES

The inequality as given by Shannon (1949, p. 63, Theorem 15) is

$$\exp\left[\frac{2}{m}H(z_1, \cdots, z_m)\right] \geqq \exp\left[\frac{2}{m}H(x_1, \cdots, x_m)\right] \\ + \exp\left[\frac{2}{m}H(y_1, \cdots, y_m)\right] \qquad (7.1)$$

where $(x_1, \cdots, x_m)$, $(y_1, \cdots, y_m)$ are independent m-dimensional random vectors and $(z_1, \cdots, z_m)$ their sum.

We give an outline of a proof by induction. For $m = 1$, (7.1) was proved in Section 6. The case $m + 1$ can be derived from the case m as follows: If we write ξ, η, ζ for $(x_1, \cdots, x_m)$, $(y_1, \cdots, y_m)$, $(z_1, \cdots, z_m)$

and x, y, z for x_{m+1}, y_{m+1}, z_{m+1} we have, in the notation of Shannon (1949, p. 54),

$$H(x_1, \cdots, x_{m+1}) = H(\xi) + H_\xi(x)$$
$$H(y_1, \cdots, y_{m+1}) = H(\eta) + H_\eta(y) \qquad (7.2)$$
$$H(z_1, \cdots, z_{m+1}) = H(\zeta) + H_\zeta(z)$$

where

$$H_\xi(x) = \int p(\xi)q(\eta)H(x \mid \xi) \, d\xi \, d\eta$$

$$H_\eta(y) = \int p(\xi)q(\eta)H(y \mid \eta) \, d\xi \, d\eta$$

$H(x \mid \xi)$, $H(y \mid \eta)$ are the entropies for the conditional probability densities of x and y for given ξ and η. We also define

$$H_{\xi,\eta}(z) = \int p(\xi)q(\eta)H(z \mid \xi,\eta) \, d\xi \, d\eta \qquad (7.3)$$

Applying the inequality for $m = 1$ to the conditional probability densities of x, y, and $z = x + y$ for given ξ, η we have

$$\exp [2H(z \mid \xi,\eta)] \geqq \exp [2H(x \mid \xi)] + \exp [2H(y \mid \eta)]$$

Substituting this in (7.3) and applying Theorem 185 of Hardy *et al.* (1952), we find

$$\exp [2H_{\xi,\eta}(z)] \geqq \exp [2H_\xi(x)) + \exp [2H_\eta(y)] \qquad (7.4)$$

By the usual methods for proving inequalities for conditional entropies one has

$$H_\zeta(z) \geqq H_{\xi,\eta}(z) \qquad (7.5)$$

So

$$\exp \left[\frac{2}{m+1} H(z_1, \cdots, z_{m+1}) \right]$$
$$\geqq \exp \left[\frac{m}{m+1} \frac{2}{m} H(\zeta) + \frac{1}{m+1} 2H_{\xi,\eta}(z) \right] \qquad (7.6)$$

Substituting into (7.6) the relation (7.4) and the assumption of induc-

tion ((7.1) for m) and using Hardy, Littlewood, Pólya, 1952, Theorem 10, we have

$$\exp\left[\frac{2}{m+1} H(z_1, \cdots, z_{m+1})\right]$$

$$\geq \exp\left[\frac{m}{m+1} \frac{2}{m} H(\xi) + \frac{1}{m+1} 2H_\xi(x)\right]$$

$$+ \exp\left\{\frac{m}{m+1} \frac{2}{m} H(\eta) + \frac{1}{m+1} 2H_\eta(y)\right\}$$

$$= \exp\left\{\left[\frac{2}{m+1} H(x_1, \cdots, x_{m+1})\right] + \exp\left[\frac{2}{m+1} H(y_1, \cdots, y_{m+1})\right]\right\}$$

8. PROOF OF (2.3) AND (2.8)

If $I_0(p)$ exists we have by (2.12)

$$I_0(p) \exp[2H(p)] = \frac{d \exp[2H(p_v)]}{dv}\bigg|_{v=0}$$

$$= 2\pi e \lim_{v \to 0} \frac{\exp[2H(p_v)] - \exp[2H(p)]}{2\pi ev}$$

From (2.11) with q a Gaussian probability density with variance v we see that the right side is greater than or equal to 1.

We can derive (2.8) from the analogue for Fourier integrals of the Hausdorff-Young inequality. For a pair of Fourier transforms (1.11) we have, if $1 < k \leq 2$ and $k' = k/k - 1$,

$$\left[\int |\psi|^{k'} dx\right]^{1/k'} \leq \left[\int |\varphi|^k du\right]^{1/k} \tag{8.1}$$

See Titchmarsh (1937, Chapter IV, Theorem 74).[5] Writing $k = 2 - 2\epsilon$, the relation (8.1) becomes

$$\left\{\int |\psi|^{2[1+\epsilon/(1-2\epsilon)]} dx\right\}^{(1-2\epsilon)/\epsilon} \leq \left\{\left[\int |\varphi|^{2(1-\epsilon)} du\right]^{-1/\epsilon}\right\}^{-1}$$

We pass to the limit for $\epsilon \to 0$ and from Hardy et al. (1952, Theorem 3) —or rather its analogue for integrals[6]—we see that (2.8) holds.

[5] Formula (4.1.2.) of Titchmarsh (1937) contains a misprint. The multiplicative constant arises from the fact that Titchmarsh uses a notation for Fourier transforms that is different from (1.11) here.

[6] The limit as $r \to 0$ of a mean of index r is the geometric mean.

In connection with this proof it is perhaps interesting to make the following remark: $\exp[-H(p)]$ is the geometric mean of p with relation to p itself, or the mean of index 0 (M_0) in the notation of Hardy *et al.* (1952, Chapter II). The means of index r, $M_r(p,p)$, have some properties that are closely related to those of $\exp[-H(p)]$; for example, properties of convexity or concavity.[6] Further one can prove that of all expressions $F(\int G(p)\,dx_1 \cdots dx_n)$ with F, G functions of one real argument, only the generalized means $M_r(p,p)$ are multiplicative in independent random variables $x_1, \cdots, x_n$.

After completion of this work and submission for publication, a referee informed me of the paper by Hirschman, which contains a proof of my Eq. (2.8) and hypothesizes Eq. (2.7), and also of the paper by Bourret (1958), which comments on the physical implications of these results.

REFERENCES

BOURRET, R. (1958). A note on an information theoretic form of the uncertainty principle. *Inform. and Control* **1**, 398.

FISHER, R. A. (1925). *Proc. Cambridge Phil. Soc.* **22**, 700.

HARDY, G. H., LITTLEWOOD, J. E., AND PÓLYA, G. (1952). "Inequalities," 2nd ed. Cambridge Univ. Press, London and New York.

HIRSCHMAN, I. I., JR. (1957). *Am. J. Math.* **79**, 152–156.

PITMAN, E. J. G. (1936). *Proc. Cambridge Phil. Soc.* **32**, 567.

SHANNON, C. E., AND WEAVER, W. (1949). "The Mathematical Theory of Communication." Univ. Illinois Press, Urbana, Illinois.

TITCHMARSH, E. C. (1937). "Introduction to the Theory of Fourier Integrals." Oxford, Univ. Press, London and New York.

WEYL, H. (1949). "The Theory of Groups and Quantum Mechanics," Appendix 1. Translation from the 2nd German edition by H. P. Robertson. Dover, New York.