Number 15 2005

RATIO MATHEMATICA

Journal of Applied Mathematics

ISSN: 1592 - 7415

The general solutions of a functional equation related to information theory

Prem Nath and Dhiraj Kumar Singh

Department of Mathematics, University of Delhi
Delhi – 110007, INDIA

E-mail: dksingh@maths.du.ac.in
dhiraj426@rediffmail.com

Abstract. The general solutions of a functional equation, containing two unknown functions, and related to a functional equation characterizing the Shannon entropy and the entropy of degree α, are obtained.

Keywords: Functional equations, continuous solutions, Lebesgue measurable solutions, the Shannon entropy, the nonadditive entropy of degree α, multiplicative functions, additive functions.

1. Introduction

For $n = 1, 2, 3, \ldots$, let

$$\Gamma_n = \left\{ (p_1, \ldots, p_n) : p_i \geq 0, i = 1, \ldots, n; \sum_{i=1}^{n} p_i = 1 \right\}$$

denote the set of all n-component complete discrete probability distributions with nonnegative elements and let $F : I \to \mathbb{R}$, $\mathbb{R}$ denoting the set of all real numbers and $I = \{x \in \mathbb{R} : 0 \leq x \leq 1\}$, the unit closed interval.

The functional equation

$$\sum_{i=1}^{k} \sum_{j=1}^{\ell} F(p_i q_j) = \sum_{i=1}^{k} F(p_i) + \sum_{j=1}^{\ell} F(q_j) \tag{1.1}$$

with $(p_1, \ldots, p_k) \in \Gamma_k$ and $(q_1, \ldots, q_\ell) \in \Gamma_\ell$ was first studied by T.W. Chaundy and J.B. Mcleod [4]. They proved that if (1.1) holds for integers $k = 2, 3, \ldots$ and $\ell = 2, 3, \ldots$ and F is continuous on I, then F is of the form

$$F(p) = c\, p \log_2 p\,, \qquad 0 \leq p \leq 1 \tag{1.2}$$

where c is an arbitrary real constant and $0 \log_2 0 = 0$. Later on, J. Aczél and Z. Daróczy [1] proved the same by assuming $k = \ell = 2, 3, \ldots$. Z. Daróczy [5] obtained the Lebesgue measurable solutions of (1.1) by fixing $k = 3$, $\ell = 2$ and assuming $F(1) = 0$. Gy. Maksa [13] obtained the solutions of (1.1) by fixing $k = 3$, $\ell = 2$ but assuming F to be bounded on a subset, of I, of positive Lebesgue measure.

If $F\left(\frac{1}{2}\right) = \frac{1}{2}$, then (1.2) gives $c = -1$ and then (1.2) reduces to

$$F(p) = -\, p \log_2 p \tag{1.3}$$

for all $p \in I$.

For any probability distribution $(p_1,\ldots,p_m)\in\Gamma_m$,

$$H_m(p_1,\ldots,p_m)=-\sum_{i=1}^{m}p_i\log_2 p_i \tag{1.4}$$

is known as the Shannon entropy [15] of the probability distribution $(p_1,\ldots,p_m)\in\Gamma_m$ and the sequence $H_m:\Gamma_m\to\mathbb{R},\ m=1,2,\ldots$ is known as the sequence of the Shannon entropies.

A generalization of the Shannon entropy with which we shall be concerned in this paper is (with $H_m^{\alpha}:\Gamma_m\to\mathbb{R},\ m=1,2,3,\ldots$)

$$H_m^{\alpha}(p_1,\ldots,p_m)=(1-2^{1-\alpha})^{-1}\left(1-\sum_{i=1}^{m}p_i^{\alpha}\right),\ \alpha>0,\alpha\neq1,0^{\alpha}:=0,\alpha\in\mathbb{R}. \tag{1.5}$$

The entropies (1.5) are due to J. Havrda and F. Charvat [7].

The axiomatic characterization of the entropies (1.5) leads to the study of the functional equation

$$\sum_{i=1}^{k}\sum_{j=1}^{\ell}F(p_iq_j)=\sum_{i=1}^{k}F(p_i)+\sum_{j=1}^{\ell}F(q_j)+\lambda\sum_{i=1}^{k}F(p_i)\sum_{j=1}^{\ell}F(q_j) \tag{1.6}$$

where $(p_1,\ldots,p_k)\in\Gamma_k$, $(q_1,\ldots,q_\ell)\in\Gamma_\ell$ and $\lambda=2^{1-\alpha}-1,\ \alpha\in\mathbb{R}$. Clearly, (1.6) reduces to (1.1) if $\lambda=0$.

By taking $\lambda=2^{1-\alpha}-1,\ \alpha\neq1,\ \alpha\in\mathbb{R},\ 0^{\alpha}:=0$, the continuous solutions of (1.6) were obtained by M. Behara and P. Nath [3] for all positive integers $k=2,3,\ldots;\ell=2,3,\ldots$. Later on PL. Kannappan [10] and D.P. Mittal [14] also obtained the continuous solutions of (1.6) for $\lambda\neq0$ and $k=2,3,\ldots;\ \ell=2,3,\ldots$. For fixed integers $k\geq3$ and $\ell\geq2$, L. Losonczi [11] obtained the measurable solutions of (1.6). Also, PL. Kannappan [8] obtained the Lebesgue measurable solutions of both (1.1) and (1.6) for fixed integers $k\geq3,\ \ell\geq3$.

It seems that L. Losonczi and Gy. Maksa [12] are the first to obtain the general solutions of (1.6) in both cases, namely $\lambda \neq 0$ and $\lambda = 0$, by fixing integers k and $\ell, k \geq 3$ and $\ell \geq 3$.

There are several generalizations of (1.6), with $\lambda \in \mathbb{R}$, containing at least two unknown functions. Below we list only three important generalizations of (1.6), namely,

$$\sum_{i=1}^{k}\sum_{j=1}^{\ell} F(p_i q_j) = \sum_{i=1}^{k} H(p_i) + \sum_{j=1}^{\ell} H(q_j) + \lambda \sum_{i=1}^{k} H(p_i) \sum_{j=1}^{\ell} H(q_j) \tag{1.7}$$

$$\sum_{i=1}^{k}\sum_{j=1}^{\ell} F(p_i q_j) = \sum_{i=1}^{k} F(p_i) + \sum_{j=1}^{\ell} H(q_j) + \lambda \sum_{i=1}^{k} F(p_i) \sum_{j=1}^{\ell} H(q_j) \tag{1.8}$$

and

$$\sum_{i=1}^{k}\sum_{j=1}^{\ell} F(p_i q_j) = \sum_{i=1}^{k} G(p_i) + \sum_{j=1}^{\ell} H(q_j) + \lambda \sum_{i=1}^{k} G(p_i) \sum_{j=1}^{\ell} H(q_j). \tag{1.9}$$

The object of this paper is to investigate the general solutions of the functional equation (1.7) for fixed integers $k \geq 3$ and $\ell \geq 3$. The corresponding results for the functional equations (1.8) and (1.9) have also been investigated by the authors and shall be presented elsewhere in our subsequent research work.

The process of finding the general solutions of (1.7) requires a detailed study of the following two functional equations :

$$\sum_{i=1}^{k}\sum_{j=1}^{\ell} g(p_i q_j) = \sum_{i=1}^{k} g(p_i) \sum_{j=1}^{\ell} g(q_j) + \ell(k-1)\, g(0) \tag{1.10}$$

and

$$\sum_{i=1}^{k}\sum_{j=1}^{\ell} f(p_i q_j) = \sum_{i=1}^{k} h(p_i) \sum_{j=1}^{\ell} h(q_j) \tag{1.11}$$

where $f : [0,1] \to \mathbb{R}$, $g : [0,1] \to \mathbb{R}$ and $h : [0,1] \to \mathbb{R}$.

The functional equation (1.10) is, indeed, a generalization of the multiplicativetype functional equation

$$\sum_{i=1}^{k}\sum_{j=1}^{\ell} g(p_i q_j) = \sum_{i=1}^{k} g(p_i)\sum_{j=1}^{\ell} g(q_j) \tag{1.12}$$

whose importance in information theory is well-known (see L. Losonczi and Gy. Maksa [12]). The functional equation (1.6), for $\lambda \neq 0$, can be written in the multiplicative form (1.12) by defining $g : I \to \mathbb{R}$ as $g(x) = \lambda F(x) + x$ for all $x \in I$. Likewise, each of the functional equations (1.7), (1.8) and (1.9), for $\lambda \neq 0$, can also be written in the corresponding multiplicative forms. This is precisely the reason for paying attention to the functional equations (1.7) to (1.9).

2. The general solutions of functional equation (1.10)

Before investigating the general solutions of (1.10) for fixed integers k and ℓ, $k \geq 3$, $\ell \geq 3$, we need some definitions and results already existing in the literature (see [12]). Let

$$\Delta = \left\{(x, y) : 0 \leq x \leq 1,\ 0 \leq y \leq 1,\ 0 \leq x + y \leq 1\right\}.$$

In other words, Δ denotes the unit closed triangle in

$$\mathbb{R}^2 = \mathbb{R} \times \mathbb{R} = \left\{(x, y) : x \in \mathbb{R}, y \in \mathbb{R}\right\}.$$

A mapping $a : \mathbb{R} \to \mathbb{R}$ is said to be additive if it satisfies the equation

$$a(x + y) = a(x) + a(y) \tag{2.1}$$

for all $x \in \mathbb{R}$, $y \in \mathbb{R}$.

A mapping $a : I \to \mathbb{R}$, $I = [0, 1]$ is said to be additive on the triangle Δ if it satisfies (2.1) for all $(x, y) \in \Delta$.

A mapping $m : [0,1] \to \mathbb{R}$ is said to be multiplicative if $m(0) = 0$, $m(1) = 1$ and $m(xy) = m(x)\, m(y)$ for all $x \in]0,1[, y \in]0,1[$.

Now we state :

Lemma 1. *Let $\Psi : I \to \mathbb{R}$ be a mapping which satisfies the functional equation*

$$\sum_{i=1}^{n} \Psi(p_i) = c \tag{2.2}$$

for all $(p_1, \ldots, p_n) \in \Gamma_n$; c a given constant and $n \geq 3$ a fixed integer. Then there exists an additive mapping $a : \mathbb{R} \to \mathbb{R}$ such that

$$\Psi(p) = a(p) + \Psi(0)\,, \qquad 0 \leq p \leq 1 \tag{2.3}$$

where

$$a(1) = c - n\,\Psi(0)\,. \tag{2.4}$$

Conversely, if (2.4) holds, then the mapping $\Psi : I \to \mathbb{R}$, defined by (2.3), satisfies the functional equation (2.2).

This lemma appears on p-74 in [12].

Lemma 2. *Every mapping $a : I \to \mathbb{R}$, $I = [0,1]$, additive on the unit triangle Δ, has a unique additive extension to the whole of $\mathbb{R}$.*

Note. This unique additive extension to the whole of $\mathbb{R}$ will also be denoted by the symbol a but now $a : \mathbb{R} \to \mathbb{R}$.

For Lemma 2, See Theorem (0.3.7) on p-8 in [2] or Z. Daróczy and L. Losonczi [6]. Now we prove :

Theorem 1. *Let $k \geq 3$, $\ell \geq 3$ be fixed integers and $g : [0,1] \to \mathbb{R}$ be a mapping which satisfies the functional equation* (1.10) *for all $(p_1, \ldots, p_k) \in \Gamma_k$ and $(q_1, \ldots, q_\ell) \in \Gamma_\ell$. Then g is of the form*

$$g(p) = a(p) + g(0) \tag{2.5}$$

where $a : \mathbb{R} \to \mathbb{R}$ is an additive function such that $a(1)$ satisfies the equation

$$a(1) + k\ell\, g(0) = [a(1) + k\, g(0)][a(1) + \ell\, g(0)] + \ell(k-1)\, g(0) \tag{2.6}$$

or

$$g(p) = M(p) - A(p) + g(0) \tag{2.7}$$

where $A : \mathbb{R} \to \mathbb{R}$ is an additive function with

$$A(1) = \ell\, g(0) \tag{2.8}$$

and $M : [0,1] \to \mathbb{R}$ is a mapping such that

$$M(0) = 0 \tag{2.9}$$

$$M(1) = g(1) + (\ell - 1)\, g(0) \tag{2.10}$$

and

$$M(pq) = M(p)\, M(q) \quad \text{for all} \quad p \in]0,1[\,,\ q \in]0,1[\,. \tag{2.11}$$

Proof. Let us put $p_1 = 1$, $p_2 = \ldots = p_k = 0$ in (1.10). We obtain

$$[1 - g(1) - (k-1)g(0)] \sum_{j=1}^{\ell} g(q_j) = 0. \tag{2.12}$$

CASE 1. $1 - g(1) - (k-1)\,g(0) \neq 0$. Then (2.12) reduces to

$$\sum_{j=1}^{\ell} g(q_j) = 0. \tag{2.13}$$

Hence, by Lemma 1, g is of the form (2.5) in which $a : \mathbb{R} \to \mathbb{R}$ is an additive mapping such that $a(1) = -\,\ell\, g(0)$ satisfies the equation (2.6).

CASE 2. $1 - g(1) - (k-1)\,g(0) = 0$.

The functional equation (1.10) may be written in the form

$$\sum_{j=1}^{\ell} \left[\sum_{i=1}^{k} g(p_i q_j) - g(q_j) \sum_{i=1}^{k} g(p_i) \right] = \ell(k-1)\, g(0).$$

Hence, by Lemma 1,

$$\begin{aligned} &\sum_{i=1}^{k} g(p_i q) - g(q) \sum_{i=1}^{k} g(p_i) \\ &\quad = \; A_1(p_1, \ldots, p_k, q) - \frac{1}{\ell}\, A_1(p_1, \ldots, p_k, 1) + (k-1)\, g(0) \end{aligned} \tag{2.14}$$

where $A_1 : \Gamma_k \times \mathbb{R} \to \mathbb{R}$ is additive in the second variable. The substitution $q = 0$ in (2.14) gives

$$A_1(p_1, \ldots, p_k, 1) = \ell\, g(0) \left[\sum_{i=1}^{k} g(p_i) - 1 \right]. \tag{2.15}$$

Let $x \in [0,1]$, $(r_1, \ldots, r_k) \in \Gamma_k$. Put $q = x r_t$, $t = 1, \ldots, k$ in (2.14); add the resulting k equations; and use the additivity of A_1. We get

$$\begin{aligned} &\sum_{i=1}^{k} \sum_{t=1}^{k} g(p_i r_t x) - \sum_{i=1}^{k} g(p_i) \sum_{t=1}^{k} g(x r_t) \\ &\quad = \; A_1(p_1, \ldots, p_k, x) - \frac{k}{\ell}\, A_1(p_1, \ldots, p_k, 1) + k(k-1)\, g(0). \end{aligned} \tag{2.16}$$

Now put $q = x,\ p_1 = r_1, \ldots, p_k = r_k$ in (2.14). We obtain

$$\sum_{t=1}^{k} g(xr_t) = g(x)\sum_{t=1}^{k} g(r_t) + A_1(r_1, \ldots, r_k, x) - \frac{1}{\ell}\, A_1(r_1, \ldots, r_k, 1) + (k-1)\, g(0). \tag{2.17}$$

From (2.16) and (2.17), it follows that

$$\begin{aligned} &\sum_{i=1}^{k}\sum_{t=1}^{k} g(p_i r_t x) - g(x)\sum_{i=1}^{k} g(p_i)\sum_{t=1}^{k} g(r_t) - k(k-1)\, g(0) \\ &= (k-1)\, g(0)\sum_{i=1}^{k} g(p_i) + A_1(r_1, \ldots, r_k, x)\sum_{i=1}^{k} g(p_i) \\ &\quad - \frac{1}{\ell}\, A_1(r_1, \ldots, r_k, 1)\sum_{i=1}^{k} g(p_i) + A_1(p_1, \ldots, p_k, x) \\ &\quad - \frac{k}{\ell}\, A_1(p_1, \ldots, p_k, 1). \end{aligned} \tag{2.18}$$

The left hand side of (2.18) does not undergo any change if we interchange p_i and $r_i, i = 1, \ldots, k$. So, the right hand side of (2.18) must also remain unchanged on interchanging p_i and $r_i,\ i = 1, \ldots, k$. Consequently, we obtain

$$\begin{aligned} &A_1(p_1, \ldots, p_k, x)\left[\sum_{t=1}^{k} g(r_t) - 1\right] - \frac{1}{\ell}\, A_1(p_1, \ldots, p_k, 1)\left[\sum_{t=1}^{k} g(r_t) - k\right] \\ &+ (k-1)\, g(0)\sum_{t=1}^{k} g(r_t) \\ &= A_1(r_1, \ldots, r_k, x)\left[\sum_{i=1}^{k} g(p_i) - 1\right] - \frac{1}{\ell}\, A_1(r_1, \ldots, r_k, 1)\left[\sum_{i=1}^{k} g(p_i) - k\right] \\ &\quad + (k-1)\, g(0)\sum_{i=1}^{k} g(p_i). \end{aligned} \tag{2.19}$$

Now we divide our discussion into two cases depending upon whether $\sum_{t=1}^{k} g(r_t) - 1$ vanishes identically on Γ_k or does not vanish identically on Γ_k.

CASE 2.1. $\sum_{t=1}^{k} g(r_t) - 1$ vanishes identically on Γ_k. Then

$$\sum_{t=1}^{k} g(r_t) = 1$$

for all $(r_1, \ldots, r_k) \in \Gamma_k$. By using Lemma 1, it follows that g is of the form (2.5) in which $a(1) = 1 - k\,g(0)$ satisfies the equation (2.6).

CASE 2.2. $\sum_{t=1}^{k} g(r_t) - 1$ does not vanish identically on Γ_k.

In this case, there exists a probability distribution $(r_1^*, \ldots, r_k^*) \in \Gamma_k$ such that

$$\sum_{t=1}^{k} g(r_t^*) - 1 \neq 0. \tag{2.20}$$

Putting $r_1 = r_1^*, \ldots, r_k = r_k^*$ in (2.19), making use of (2.20) and (2.15); and performing necessary calculations, it follows that

$$A_1(p_1, \ldots, p_k, x) = A(x) \left[\sum_{i=1}^{k} g(p_i) - 1 \right] \tag{2.21}$$

where $A : \mathbb{R} \to \mathbb{R}$ is such that

$$A(x) = \left[\sum_{t=1}^{k} g(r_t^*) - 1 \right]^{-1} A_1(r_1^*, \ldots, r_k^*, x) \tag{2.22}$$

From (2.22) it is easy to conclude that $A : \mathbb{R} \to \mathbb{R}$ is additive as the mapping $x \mapsto A_1(r_1^*, \ldots, r_k^*, x)$ is additive. Also, putting $x = 1$ in (2.22) and making use of (2.15) by taking $p_i = r_i^*$, $i = 1, \ldots, k$; (2.8) follows. Also, from (2.14), (2.15), (2.21) and (2.8), it follows that

$$\begin{aligned} &\sum_{i=1}^{k} [g(p_i q) + A(p_i q) - g(0)] - [g(q) + A(q) - g(0)] \\ &\qquad \times \sum_{i=1}^{k} [g(p_i) + A(p_i) - g(0)] + [g(q) + A(q) - g(0)](\ell - k)\, g(0) = 0. \end{aligned} \tag{2.23}$$

Define a mapping $M : I \to \mathbb{R},\ I = [0,1]$, as

$$M(p) = g(p) + A(p) - g(0) \tag{2.24}$$

for all $p \in I$. Then, (2.23) reduces to the equation

$$\sum_{i=1}^{k} [M(p_i q) - M(q)\, M(p_i) + (\ell - k)\, g(0)\, M(q)\, p_i] = 0. \tag{2.25}$$

Hence, by Lemma 1,

$$M(pq) - M(q)\, M(p) + (\ell - k)\, g(0)\, M(q)\, p = E_1(p,q) - \frac{1}{k}\, E_1(1,q) \tag{2.26}$$

where $E_1 : \mathbb{R} \times [0,1] \to \mathbb{R}$ is additive in its first variable.

Since $A(0) = 0$ and $A(1) = \ell\, g(0)$, (2.9) and (2.10) follow from (2.24). Also, putting $p = 0$ in (2.26) and making use of (2.9), it follows that

$$E_1(0,q) = 0 \tag{2.27}$$

for all $q,\ 0 \le q \le 1$. Consequently,

$$E_1(1,q) = 0 \tag{2.28}$$

for all $q,\ 0 \le q \le 1$. Now, (2.26) reduces to

$$M(pq) - M(p)\, M(q) = E_1(p,q) - (\ell - k)\, g(0)\, M(q)\, p \tag{2.29}$$

for all $p \in [0,1]$ and $q \in [0,1]$.

Since $M(1) = g(1) + (\ell - 1)\, g(0)$, from now onwards, we divide our discussion into two subcases, depending upon whether $g(1) + (\ell - 1)\, g(0) = 1$ or $g(1) + (\ell - 1)\, g(0) \neq 1$.

Case 2.2.1. $g(1) + (\ell - 1)\, g(0) = 1$.

In this case, $1 = g(1) + (\ell - 1)\, g(0) = g(1) + (k - 1)\, g(0) + (\ell - k)\, g(0)$.

Since $g(1)+(k-1)g(0)=1$, it follows that $(\ell-k)\,g(0)=0$. Then, (2.29) reduces to

$$M(pq)-M(p)\,M(q)=E_1(p,q) \tag{2.30}$$

where $E_1:\mathbb{R}\times[0,1]\to\mathbb{R}$ is additive in the first variable and $0\le p\le 1$, $0\le q\le 1$. The left hand side of (2.30) is symmetric in p and q. Hence, $E_1(p,q)=E_1(q,p)$ for all $p\in[0,1], q\in[0,1]$. Consequently, E_1 is also additive in second variable. Also, we may suppose that $E_1(p,\cdot)$ has been extended additively to the whole of $\mathbb{R}$ and this extension is unique by Lemma 2.

From (2.30), as on p-77 in [12], it follows that

$$\begin{aligned} M(pqr)-M(p)\,M(q)\,M(r) &= E_1(pq,r)+M(r)\,E_1(p,q) \\ &= E_1(qr,p)+M(p)\,E_1(q,r) \end{aligned} \tag{2.30a}$$

for all $p,\ q,\ r$ in [0,1]. Now, we prove that $E_1(p,q)=0$ for all $p,\ q,\ 0\le p\le 1$, $0\le q\le 1$. If possible, suppose there exist p^* and q^*, $0\le p^*\le 1,\ 0\le q^*\le 1$, such that $E_1(p^*,q^*)\neq 0$. Then, from (2.30a),

$$M(r)=\big[E_1(p^*,q^*)\big]^{-1}\big[E_1(q^*r,p^*)+M(p^*)\,E_1(q^*,r)-E_1(p^*q^*,r)\big]$$

from which it is easy to conclude that M is additive. Now, making use of (2.8), (2.10), (2.20), (2.24), the condition $g(1)+(\ell-1)\,g(0)=1$; and the additivity of A and M, we have

$$1\neq\sum_{t=1}^{k}g(r_t^*)=M(1)-A(1)+k\,g(0)=1$$

a contradiction. Hence $E_1(p,q)=0$ for all p and q, $0\le p\le 1$, $0\le q\le 1$. Thus, (2.30) reduces to $M(pq)=M(p)\,M(q)$ for all p and q, $0\le p\le 1$, $0\le q\le 1$. So, M is a nonconstant multiplicative function. Hence, from (2.24), it follows that g is of the form (2.7).

CASE 2.2.2. $g(1)+(\ell-1)\,g(0)\neq 1$.

Since the values of M at 0 and 1 are given by (2.9) and (2.10), our next task is to get some information about $M(r)$ when $0<r<1$. For this purpose, we proceed as follows:

Let $p,\ q,\ r$ be in]0,1[. Now, from (2.29), one can derive

$$\begin{aligned}(2.31)\quad & M(pqr)-M(p)\,M(q)\,M(r)\\ &= E_1(r,pq)-(\ell-k)\,g(0)\,M(pq)\,r+M(r)\big[E_1(p,q)-(\ell-k)\,g(0)\,M(q)\,p\big]\\ &= E_1(rq,p)-(\ell-k)\,g(0)\,M(p)\,rq+M(p)\big[E_1(r,q)-(\ell-k)\,g(0)\,M(q)\,r\big].\end{aligned}$$

Now, we prove that $E_1(p,q)-(\ell-k)\,g(0)\,M(q)\,p=0$ for all $p,\ q,\ 0<p<1$, $0<q<1$. If possible, suppose there exist $p^*\in\,]0,1[$ and $q^*\in\,]0,1[$ such that $E_1(p^*,q^*)-(\ell-k)\,g(0)\,M(q^*)\,p^*\neq 0$. Then, from (2.31), it follows that for all $r\in\,]0,1[$,

$$\begin{aligned}(2.32)\quad M(r) \;=\;& \big[E_1(p^*,q^*)-(\ell-k)\,g(0)\,M(q^*)\,p^*\big]^{-1}\\ &\times\big[E_1(rq^*,p^*)-(\ell-k)\,g(0)\,M(p^*)\,rq^*\\ &+M(p^*)\big\{E_1(r,q^*)-(\ell-k)\,g(0)\,M(q^*)\,r\big\}-E_1(r,p^*q^*)\\ &+(\ell-k)\,g(0)\,M(p^*q^*)\,r\big].\end{aligned}$$

Now we prove that $M:[0,1]\to\mathbb{R}$ is additive on Δ, that is,

$$(2.33)\quad M(x+y)=M(x)+M(y)$$

for all $0\le x\le 1,\ 0\le y\le 1,\ 0\le x+y\le 1$.

If $x=0,\ 0\le y\le 1$ or $y=0,\ 0\le x\le 1$, then (2.33) holds trivially.

If $0<x<1,\ 0<y<1,\ 0<x+y<1$, then (2.33) follows from (2.32).

Now consider the case when $0<x<1,\ 0<y<1$ but $x+y=1$. In this case, let us choose $q=1$ and $p=x+y$ in (2.29) and use the additivity of E_1

with respect to first variable. We obtain

$$\begin{aligned} & M(x+y)\big\{1-g(1)-(\ell-1)\,g(0)\big\} \\ &= \big\{M(x)+M(y)\big\}\big\{1-g(1)-(\ell-1)\,g(0)\big\}. \end{aligned}$$

Since $g(1)+(\ell-1)\,g(0)\neq 1$, (2.33) follows. Thus, M is additive on the triangle Δ. Now, making use of (2.8), (2.10), (2.20), (2.24), the condition $g(1)+(k-1)\,g(0)=1$; and the additivity of A and M, we have

$$1\neq\sum_{t=1}^{k} g(r_t^*)=M(1)-A(1)+k\,g(0)=1$$

a contradiction. Hence $E_1(p,q)-(\ell-k)\,g(0)\,M(q)\,p=0$ for all $p,\ q, 0<p<1$, $0<q<1$. Thus, (2.29) reduces to (2.11). But in this case M is not multiplicative because $M(1)=g(1)+(\ell-1)\,g(0)\neq 1$. Hence from (2.24), the solution (2.7) follows.

Note. It is easy to verify that (2.5); subject to the condition (2.6), satisfies (1.10). However (2.7) also satisfies (1.10). In this case we need to use (2.25) in addition to (2.8) to (2.11).

3. The general solutions of functional equation (1.11)

Now we prove :

Theorem 2. *Let $k\geq 3$, $\ell\geq 3$ be fixed integers and $f: I\to\mathbb{R}$, $h: I\to\mathbb{R}$, $I=[0,1]$, be mappings which satisfy the functional equation* (1.11) *for all $(p_1,\ldots,p_k)\in\Gamma_k$ and $(q_1,\ldots,q_\ell)\in\Gamma_\ell$. Then any general solution of* (1.11) *is of the form*

$$\begin{cases} f(p)=b(p)+f(0) \\ h(p)=B(p)+h(0) \end{cases} \tag{3.1}$$

subject to the condition

$$b(1) + k\ell\, f(0) = [B(1) + k\, h(0)][B(1) + \ell\, h(0)]; \tag{3.2}$$

or

$$\begin{cases} f(p) = [h(1) + (k-1)\, h(0)]^2\, a(p) + A^*(p) + f(0) \\ h(p) = [h(1) + (k-1)\, h(0)]\, a(p) + h(0) \end{cases} \tag{3.3}$$

subject to the condition

$$\begin{aligned} [h(1) + (k-1)h(0)]^2 a(1) + A^*(1) + k\ell\, f(0) \\ = \big\{[h(1)+(k-1)h(0)]a(1)+k\, h(0)\big\}\big\{[h(1)+(k-1)h(0)]a(1)+\ell\, h(0)\big\} \end{aligned} \tag{3.3a}$$

or

$$\begin{cases} f(p) = [h(1) + (k-1)\, h(0)]^2 [M(p) - A(p)] + A^*(p) + f(0) \\ h(p) = [h(1) + (k-1)\, h(0)][M(p) - A(p)] + h(0) \\ A(1) = \dfrac{\ell\, h(0)}{h(1) + (k-1)\, h(0)}, \\ A^*(1) = \ell\, \big\{[h(1) + (k-1)\, h(0)]\, h(0) - k\, f(0)\big\} \end{cases} \tag{3.4}$$

where $A^* : \mathbb{R} \to \mathbb{R}$, $A : \mathbb{R} \to \mathbb{R}$, $B : \mathbb{R} \to \mathbb{R}$, $a : \mathbb{R} \to \mathbb{R}$, $b : \mathbb{R} \to \mathbb{R}$ *are additive functions;* $f(0)$ *and* $h(0)$ *are arbitrary constants; and* $M : [0, 1] \to \mathbb{R}$ *is a mapping which satisfies* (2.9), (2.11) *and*

$$M(1) = \frac{h(1) + (\ell - 1)\, h(0)}{h(1) + (k-1)\, h(0)} \tag{3.5}$$

with $h(1) + (k-1)\, h(0) \neq 0$ *in* (3.3), (3.3a), (3.4) *and* (3.5).

To prove this theorem, we need to prove some Lemmas :

Lemma 3. *If a mapping* $f : I \to \mathbb{R}$ *satisfies the functional equation*

$$\sum_{i=1}^{k} \sum_{j=1}^{\ell} f(p_i q_j) = 0 \tag{3.6}$$

for all $(p_1, \ldots, p_k) \in \Gamma_k$, $(q_1, \ldots, q_\ell) \in \Gamma_\ell$, $k \geq 3, \ell \geq 3$ *fixed integers; then*

$$f(p) = b(p) + f(0) \tag{3.7}$$

where $b : \mathbb{R} \to \mathbb{R}$ *is an additive function with* $b(1) = -k\ell\, f(0)$.

Proof. Choose $q_1 = 1$, $q_2 = \ldots = q_\ell = 0$. Then, equation (3.6) reduces to

$$\sum_{i=1}^{k} f(p_i) = -k(\ell - 1)\, f(0).$$

Hence, by Lemma 1,

$$f(p) = b(p) - \frac{1}{k}\, b(1) - (\ell - 1)\, f(0) \tag{3.8}$$

for all p, $0 \leq p \leq 1$, $b : \mathbb{R} \to \mathbb{R}$ being any additive function with $b(1) = -k\ell\, f(0)$. Putting this value of $b(1)$ in (3.8), (3.7) readily follows.

Lemma 4. *Under the assumptions stated in the statement of* Theorem 2*, the following conclusions hold :*

$$f(p) = [h(1)+(k-1)h(0)]h(p)+A^*(p)-[h(1)+(k-1)h(0)]h(0)+f(0) \tag{3.9}$$

$$[h(1) + (k-1)\, h(0)] \sum_{i=1}^{k} \sum_{j=1}^{\ell} h(p_i q_j) - \sum_{i=1}^{k} h(p_i) \sum_{j=1}^{\ell} h(q_j) \tag{3.10}$$

$$= \ell(k-1)\, h(0)[h(1) + (k-1)\, h(0)]$$

$$[h(1) + (\ell-1)\, h(0)] \sum_{i=1}^{k} \sum_{j=1}^{\ell} h(p_i q_j) - \sum_{i=1}^{k} h(p_i) \sum_{j=1}^{\ell} h(q_j) \tag{3.11}$$

$$= k(\ell-1)\, h(0)[h(1) + (\ell-1)\, h(0)]$$

where $A^* : \mathbb{R} \to \mathbb{R}$ *is an additive function.*

Proof. Putting $p_1 = 1,\ p_2 = \ldots = p_k = 0$ in (1.11), we obtain

$$\sum_{j=1}^{\ell} \big\{ f(q_j) - [h(1) + (k-1)\, h(0)]\, h(q_j) \big\} = -\,\ell(k-1)\, f(0). \tag{3.12}$$

Hence, by Lemma 1 (changing q to p),

$$f(p) = [h(1) + (k-1)\, h(0)]\, h(p) + A^*(p) - \frac{1}{\ell}\, A^*(1) - (k-1)\, f(0) \tag{3.13}$$

for all $p,\ 0 \le p \le 1,\ A^* : \mathbb{R} \to \mathbb{R}$ being an additive function with

$$A^*(1) = \ell \big\{ [h(1) + (k-1)\, h(0)]\, h(0) - k\, f(0) \big\}. \tag{3.14}$$

From equations (3.13) and (3.14), equation (3.9) follows.

From (3.9) and (3.14), it is easy to see that

$$\begin{aligned} \sum_{i=1}^{k} \sum_{j=1}^{\ell} f(p_i q_j) \;\; &= \;\; [h(1) + (k-1)\, h(0)] \sum_{i=1}^{k} \sum_{j=1}^{\ell} h(p_i q_j) \\ &\quad - \ell(k-1)\, [h(1) + (k-1)\, h(0)]\, h(0). \end{aligned} \tag{3.15}$$

From (1.11) and (3.15), we get (3.10). The proof of (3.11) is similar and hence omitted.

Proof of Theorem 2. We divide our discussion into three cases :

Case 1. $\sum\limits_{i=1}^{k} h(p_i)$ vanishes identically on Γ_k, that is,

$$\sum_{i=1}^{k} h(p_i) = 0 \tag{3.16}$$

for all $(p_1, \ldots, p_k) \in \Gamma_k$. Then, (1.11) reduces to (3.6). So, f is of the form (3.7) for all $p,\ 0 \le p \le 1$. Also applying Lemma 1 to (3.16), we obtain

$$h(p) = B(p) - \frac{1}{k}\, B(1) \tag{3.17}$$

for all $p,\ 0 \le p \le 1,\ B : \mathbb{R} \to \mathbb{R}$ being an additive function with $B(1) = -\,k\, h(0)$.

Now (3.17) reduces to

$$h(p) = B(p) + h(0). \tag{3.18}$$

Equations (3.7), (3.18), together with the condition (3.2), constitute the solution (3.1) of (1.11).

Case 2. $\sum_{j=1}^{\ell} h(q_j)$ vanishes identically on Γ_ℓ. In this case, we also get the solution (3.1), subject to the condition (3.2); of (1.11). The proof is omitted as it is similar to that in case 1.

Case 3. Neither $\sum_{i=1}^{k} h(p_i)$ vanishes identically on Γ_k nor $\sum_{j=1}^{\ell} h(q_j)$ vanishes identically on Γ_ℓ. Then, there exist a $(p_1^*, \ldots, p_k^*) \in \Gamma_k$ and a $(q_1^*, \ldots, q_\ell^*) \in \Gamma_\ell$ such that $\sum_{i=1}^{k} h(p_i^*) \neq 0$ and $\sum_{j=1}^{\ell} h(q_j^*) \neq 0$; and consequently

$$\sum_{i=1}^{k} h(p_i^*) \sum_{j=1}^{\ell} h(q_j^*) \neq 0\,. \tag{3.19}$$

Now, we prove that $h(1) + (k-1)\,h(0) \neq 0$. If possible, suppose $h(1) + (k-1)\,h(0) = 0$. Then (3.10) reduces to the equation

$$\sum_{i=1}^{k} h(p_i) \sum_{j=1}^{\ell} h(q_j) = 0$$

valid for all $(p_1, \ldots, p_k) \in \Gamma_k$ and $(q_1, \ldots, q_\ell) \in \Gamma_\ell$. In particular, $\sum_{i=1}^{k} h(p_i^*) \sum_{j=1}^{\ell} h(q_j^*) = 0$ contradicting (3.19). Hence $h(1) + (k-1)\,h(0) \neq 0$.

Similarly, making use of(3.11), we can prove that $h(1) + (\ell - 1)\,h(0) \neq 0$.

Let us consider the case when $h(1) + (k-1)\,h(0) \neq 0$. In this case, let us define a mapping $g : [0,1] \to \mathbb{R}$ as

$$g(x) = [h(1) + (k-1)\,h(0)]^{-1} h(x) \tag{3.20}$$

for all $x \in [0,1]$. Then, with the aid of (3.20), (3.10) reduces to the functional

equation (1.10). Also, from (3.20), it is easy to see that $g(1) + (k-1)\,g(0) = 1$. Consequently, from the discussion, carried out under this case, in the proof of theorem 1, it follows that g is of the form (2.5), subject to the condition (2.6); and (2.7). From equations (2.5), (2.7), (3.9) and (3.20), the solutions (3.3) subject to the condition (3.3a); and (3.4) of functional equation (1.11) follow. The details are omitted for the sake of brevity.

4. The general solutions of functional equation (1.7) when $\lambda \neq 0$

In this section we prove the following :

Theorem 3. *Let $k \geq 3$, $\ell \geq 3$ be fixed integers and $F : I \to \mathbb{R}$, $H : I \to \mathbb{R}$, $I = [0,1]$, be mappings which satisfy the functional equation* (1.7) *for all $(p_1, \ldots, p_k) \in \Gamma_k$ and $(q_1, \ldots, q_\ell) \in \Gamma_\ell$. Then, any general solution of* (1.7) *is of the form*

$$F(p) = \frac{b(p) + \lambda\, F(0) - p}{\lambda}, \quad H(p) = \frac{B(p) + \lambda\, H(0) - p}{\lambda} \tag{4.1}$$

subject to the condition

$$b(1) + \lambda k\ell\, F(0) = [B(1) + \lambda k\, H(0)][B(1) + \lambda\ell\, H(0)] \tag{4.2}$$

or

$$\begin{cases} F(p) = \dfrac{[\lambda\,(H(1) + (k-1)\,H(0)) + 1]^2\, a(p) + A^*(p) + \lambda\, F(0) - p}{\lambda} \\ H(p) = \dfrac{[\lambda\,(H(1) + (k-1)\,H(0)) + 1]\, a(p) + \lambda\, H(0) - p}{\lambda} \end{cases} \tag{4.3}$$

subject to the condition

$$\begin{aligned} [\lambda\,(H(1) + (k-1)\,H(0)) + 1]^2\, a(1) + A^*(1) + \lambda k\ell\, F(0) \\ = \big\{[\lambda\,(H(1) + (k-1)\,H(0)) + 1]\, a(1) + \lambda k\, H(0)\big\} \\ \times \big\{[\lambda\,(H(1) + (k-1)\,H(0)) + 1]\, a(1) + \lambda\ell\, H(0)\big\} \end{aligned} \tag{4.3a}$$

or

$$
\begin{cases}
F(p)=\dfrac{\left(\begin{array}{l}[\lambda\,(H(1)+(k-1)\,H(0))+1]^2[M(p)-A(p)]\\ +A^*(p)+\lambda\,F(0)-p\end{array}\right)}{\lambda} \\[2ex]
H(p)=\dfrac{[\lambda(H(1)+(k-1)\,H(0))+1][M(p)-A(p)]+\lambda\,H(0)-p}{\lambda} \\[2ex]
A(1)=\dfrac{\lambda\ell\,H(0)}{[\lambda\,(H(1)+(k-1)\,H(0))]}, \\[2ex]
A^*(1)=\lambda\ell\left\{[\lambda\,(H(1)+(k-1)\,H(0))+1]\,H(0)-k\,F(0)\right\}
\end{cases}
\tag{4.4}
$$

where $A^*:\mathbb{R}\to\mathbb{R}$, $A:\mathbb{R}\to\mathbb{R}$, $B:\mathbb{R}\to\mathbb{R}$, $a:\mathbb{R}\to\mathbb{R}$, $b:\mathbb{R}\to\mathbb{R}$ *are additive functions;* $M:[0,1]\to\mathbb{R}$ *satisfies* (2.9), (2.11) *and*

$$
M(1)=\frac{\lambda\,(H(1)+(\ell-1)\,H(0))+1}{\lambda\,(H(1)+(k-1)\,H(0))+1} \tag{4.5}
$$

with $[\lambda\,(H(1)+(k-1)\,H(0))+1]\neq 0$ *in* (4.3), (4.3a), (4.4) *and* (4.5).

Proof. Let us write (1.7) in the multiplicative form

$$
\sum_{i=1}^{k}\sum_{j=1}^{\ell}[\lambda\,F(p_iq_j)+p_iq_j]=\sum_{i=1}^{k}[\lambda\,H(p_i)+p_i]\sum_{j=1}^{\ell}[\lambda\,H(q_j)+q_j]. \tag{4.6}
$$

Define the mappings $f:I\to\mathbb{R}$, $h:I\to\mathbb{R}$ as

$$
f(x)=\lambda\,F(x)+x,\qquad h(x)=\lambda\,H(x)+x \tag{4.7}
$$

for all $x\in I$. Then, (4.6) reduces to the functional equation (1.11) whose solutions are given by (3.1) subject to the condition (3.2); (3.3) subject to (3.3a); and (3.4) in which $A^*:\mathbb{R}\to\mathbb{R}$, $A:\mathbb{R}\to\mathbb{R}$, $B:\mathbb{R}\to\mathbb{R}$, $a:\mathbb{R}\to\mathbb{R}$, $b:\mathbb{R}\to\mathbb{R}$ are additive functions; and $M:[0,1]\to\mathbb{R}$ is a mapping which satisfies (2.9), (2.11) and (3.5). Now making use of (4.7) and (3.1) subject to the condition (3.2); (3.3) subject to the condition (3.3a); and (3.4); the required solutions (4.1)

subject to the condition (4.2); (4.3) subject to the condition (4.3a) and (4.4) follow. The details are omitted.

5. The general solutions of functional equation (1.7) when $\lambda = 0$

If $\lambda = 0$, then (1.7) reduces to the functional equation

$$\sum_{i=1}^{k}\sum_{j=1}^{\ell} F(p_i q_j) = \sum_{i=1}^{k} H(p_i) + \sum_{j=1}^{\ell} H(q_j) \tag{5.1}$$

where $k \geq 3,\ \ell \geq 3$ are fixed integers and $(p_1, \ldots, p_k) \in \Gamma_k,\ (q_1, \ldots, q_\ell) \in \Gamma_\ell$.

The substitutions $p_1 = 1,\ p_2 = \ldots = p_k = 0$ in (5.1) yield

$$\sum_{j=1}^{\ell}[F(q_j) - H(q_j)] = H(1) + (k-1)\,H(0) - \ell(k-1)\,F(0). \tag{5.2}$$

Hence, by Lemma 1,

$$F(p) = H(p) + A_1^*(p) - \frac{1}{\ell}\,A_1^*(1) + \frac{1}{\ell}\big\{H(1) + (k-1)H(0) - \ell(k-1)F(0)\big\} \tag{5.3}$$

where $A_1^* : \mathbb{R} \to \mathbb{R}$ is additive with

$$A_1^*(1) = H(1) + (k + \ell - 1)\,H(0) - k\ell\,F(0). \tag{5.4}$$

From (5.3) and (5.4), we obtain

$$\sum_{i=1}^{k}\sum_{j=1}^{\ell} F(p_i q_j) = \sum_{i=1}^{k}\sum_{j=1}^{\ell} H(p_i q_j) + H(1) - (k-1)(\ell-1)\,H(0). \tag{5.5}$$

From (5.1) and (5.5), we obtain

$$\sum_{i=1}^{k}\sum_{j=1}^{\ell} H(p_i q_j) = \sum_{i=1}^{k} H(p_i) + \sum_{j=1}^{\ell} H(q_j) - \big\{H(1) - (k-1)(\ell-1)H(0)\big\}. \tag{5.6}$$

Define $H_1 : [0,1] \to \mathbb{R}$ as

$$H_1(x) = H(x) - \big\{H(1) - (k-1)(\ell-1)\,H(0)\big\}\,x \tag{5.7}$$

for all $x \in [0,1]$. Then, equation (5.6) reduces to

$$\sum_{i=1}^{k}\sum_{j=1}^{\ell} H_1(p_i q_j) = \sum_{i=1}^{k} H_1(p_i) + \sum_{j=1}^{\ell} H_1(q_j). \tag{5.8}$$

Putting $p_1 = q_1 = 1$ and $p_2 = \ldots = p_k = q_2 = \ldots = q_\ell = 0$ in (5.8), we obtain $H_1(1) = (k-1)(\ell-1)\, H_1(0)$. Define $H_2 : [0,1] \to \mathbb{R}$ as

$$H_2(x) = H_1(x) - H_1(0) - [H_1(1) - H_1(0)]\, x \tag{5.9}$$

for all $x \in [0,1]$. Then

$$\sum_{i=1}^{k}\sum_{j=1}^{\ell} H_2(p_i q_j) = \sum_{i=1}^{k} H_2(p_i) + \sum_{j=1}^{\ell} H_2(q_j) \tag{5.10}$$

where $H_2(1) = H_2(0) = 0$, and $(p_1, \ldots, p_k) \in \Gamma_k$, $(q_1, \ldots, q_\ell) \in \Gamma_\ell$, $k \geq 3$, $\ell \geq 3$ fixed integers. Theorem 2 (p-78 in [12]) may now be written as :

Theorem 4. *Let $k \geq 3$, $\ell \geq 3$ be fixed integers. The mapping $H_2 : [0,1] \to \mathbb{R}$ with $H_2(1) = 0$, $H_2(0) = 0$, defined in* (5.9) *is a solution of* (5.10) *if and only if*

$$H_2(p) = \begin{cases} a(p) + D(p,p) & \text{if } \ 0 < p \leq 1 \\ 0 & \text{if } \ p = 0 \end{cases} \tag{5.11}$$

where $a : \mathbb{R} \to \mathbb{R}$ is additive; $D : \mathbb{R} \times]0,1] \to \mathbb{R}$ is additive in the first variable and there exists a function $E : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, additive in both variables such that $E(1,1) = a(1)$ and, moreover,

$$D(pq,pq) - D(pq,p) - D(pq,q) = E(p,q) \quad \text{if } \ 0 < p \leq 1, \ 0 < q \leq 1. \tag{5.12}$$

Making use of corollary 3 on p-81 in [12], it follows that

$$H_1(p) = \begin{cases} c + c\,(k\ell - k - \ell)\, p + a(p) + D(p,p) & \text{if } \ 0 < p \leq 1 \\ c & \text{if } p = 0 \end{cases} \tag{5.13}$$

where $c = H_1(0)$ is an arbitrary real constant, $a : \mathbb{R} \to \mathbb{R}$ is additive, $D : \mathbb{R} \times]0,1] \to \mathbb{R}$ is as described above in Theorem 4. Now from (5.7) and

(5.13), we obtain

$$H(p) = \begin{cases} c\,(1-p) + d_1\,p + a(p) + D(p,p) & \text{if } \ 0 < p \leq 1 \\ c & \text{if } \ p = 0 \end{cases} \tag{5.14}$$

where $c = H(0)$ and $d_1 = H(1)$ are arbitrary real constants, $a : \mathbb{R} \to \mathbb{R}$ is additive function; $D : \mathbb{R} \times]0,1] \to \mathbb{R}$ as described above in Theorem 4. Now from (5.3), (5.4) and (5.14), we obtain

$$F(p) = \begin{cases} d_0 + d_1\,p - c\,p + a(p) + A_1^*(p) + D(p,p) & \text{if } \ 0 < p \leq 1 \\ d_0 & \text{if } \ p = 0 \end{cases} \tag{5.15}$$

where $c = H(0)$, $d_0 = F(0)$, $d_1 = H(1)$ are arbitrary real constants; $a : \mathbb{R} \to \mathbb{R}$, $A_1^* : \mathbb{R} \to \mathbb{R}$ are additive functions with $A_1^*(1)$ given by (5.4); $D : \mathbb{R} \times]0,1] \to \mathbb{R}$ as described above in Theorem 4. Thus, we have proved the following:

Theorem 5. *Let $k \geq 3$, $\ell \geq 3$ be fixed integers. The mappings $F : [0,1] \to \mathbb{R}$, $H : [0,1] \to \mathbb{R}$ satisfy the equation* (5.1) *if and only if F and H are respectively of the forms* (5.15) *and* (5.14) *with $A_1^*(1)$ given by* (5.4) *and D as described above in* Theorem 4.

References

[1] J. Aczél and Z. Daróczy, Characterisierung der Entropien positiver ordnung und der Shannonschen entropie, *Acta Math. Acad. Sci. Hungar.*, 14 (1963), 95–121.

[2] J. Aczél and Z. Daróczy, *On measures of information and their characterizations*, Academic Press, New York-San Francisco-London, 1975.

[3] M. Behara and P. Nath, *Additive and non-additive entropies of finite measurable partitions, Probability and Information Theory II*, *Lecture Notes in Math.*, Vol. 296, Berlin. Heidelberg-New York, 1973, 102–138.

[4] T.W. Chaundy and J.B. Mcleod, On a functional equation, *Edinburgh Math. Notes*, 43 (1960), 7–8.

[5] Z. Daróczy, On the measurable solutions of a functional equation, *Acta Math. Acad. Sci. Hungar.*, 22 (1971), 11–14.

[6] Z. Daróczy and L. Losonczi, Über die Erweiterung der auf einer Punktmenge additiven Funktionen, *Publ. Math. (Debrecen)*, 14 (1967), 239–245.

[7] J. Havrda and F. Charvat, Quantification method of classification process, Concept of structural α-entropy, *Kybernetika (Prague)*, 3 (1967), 30–35.

[8] PL. Kannappan, On some functional equations from additive and non-additive measures-I, *Proc. of the Edin. Mathematical Society*, 23 (1980), 145–150.

[9] PL. Kannappan, *On some functional equations from additive and non-additive measures-II, Advances in communication*; Second Internat. Conf. on Information Sciences and System (University of Patras, Patras, 1979), 1, 45–50, Reidel, Dordrecht-Boston-Mass., 1980.

[10] PL. Kannappan, On a generalization of some measures in information theory, *Glasnik Mat.*, 9 (29) (1974), 81–93.

[11] L. Losonczi, A characterization of entropies of degree α, *Metrika*, 28 (1981), 237–244.

[12] L. Losonczi and Gy. Maksa, On some functional equations of the information theory, *Acta Math. Acad. Sci. Hung.*, 39 (1982), 73–82.

[13] Gy. Maksa, On the bounded solutions of a functional equation, *Acta Math. Acad. Sci. Hung.*, 37 (1981), 445–450.

[14] D.P. Mittal, On continuous solutions of a functional equation, *Metrika*, 22 (1970), 31–40.

[15] C.E. Shannon, A mathematical theory of communication, *Bell Syst. Tech. Jour.*, 27 (1948), 378–423, 623–656.

A HYPEROPERATION DEFINED ON A GROUPOID EQUIPPED WITH A MAP

Thomas Vougiouklis
Democritus University of Thrace, School of Education
681 00 Alexandroupolis, Greece
e-mail: tvougiou@eled.duth.gr

ABSTRACT

The H_v-*structures* are hyperstructures where the equality is replaced by the non-empty intersection. The fact that this class of the hyperstructures is very large, one can use it in order to define several objects that they are not possible to be defined in the classical hypergroup theory. In the present paper we introduce a kind of hyperoperations which are defined on a set equipped with an operation or a hyperoperation and a map on itself.

AMS Subject Classification: 20N20
Key words: hyperstructures, H_v-structures.

1. Introduction

The object of this paper is the hyperstructures called H_v-*structures* introduced in 1990 [5], which satisfy the *weak axioms* where the non-empty intersection replaces the equality.

Recall some basic definitions:

Definitions 1. In a set H equipped with a hyperoperation $\cdot: H\times H \to \boldsymbol{P}(H)$, we abbreviate by *WASS* the *weak associativity*: $(xy)z \cap x(yz) \neq \emptyset$, $\forall x,y,z \in H$ and by *COW* the *weak commutativity*: $xy \cap yx \neq \emptyset$, $\forall x,y \in H$. The hyperstructure $(H,\cdot)$ is called H_v-*semigroup* if it is WASS, is called H_v-*group* if it is reproductive H_v-semigroup. The hyperstructure $(R,+,\cdot)$ is called H_v-*ring* if $(+)$ and $(\cdot)$ are WASS, the reproduction axiom is valid for $(+)$ and $(\cdot)$ is *weak distributive* with respect to $(+)$: $x(y+z) \cap (xy+xz) \neq \emptyset$, $(x+y)z \cap (xz+yz) \neq \emptyset$,

$\forall x,y,z \in R$. *H_v-modulus* and *H_v-vector spaces* are also defined in a similar way.

For more definitions, results and applications on H_v-structures, see books [6,2] and on some papers such as [3-11]. A special class [6]: An H_v-structure is called *very thin* iff all its hyperoperations are operations except one, which all hyperproducts are singletons except only one, which has cardinality more than one.

The fundamental relations β^*, γ^* and ε^* are defined, in H_v-groups, H_v-rings and H_v-vector spaces, respectively, as the smallest equivalences so that the quotient would be group, ring and vector space, respectively (see [1,6]). The way to find the fundamental classes is given by analogous theorems to the following [5,6,7]:

Theorem. Let $(H,\cdot)$ be H_v-group and let us denote by $\boldsymbol{U}$ the set of all finite products of elements of H. We define the relation β in H as follows: $x\beta y$ iff $\{x,y\} \subset \boldsymbol{u}$ where $\boldsymbol{u} \in \boldsymbol{U}$. Then the fundamental relation β^* is the transitive closure of β.

Proof. The main point is: Take x,y such that $\{x,y\} \subset \boldsymbol{u} \in \boldsymbol{U}$ and any hyperproduct where one of the elements x,y, is used. Then, if this element is replaced by the other, the new hyperproduct is inside the same fundamental class where the first hyperproduct is. Therefore, if the hyperproducts of the above β-classes are products, then, they are fundamental classes. Analogous remarks for the relations γ^*, in H_v-rings, and ε^*, in H_v-vector spaces, are also applied.

An element is called *single* if its fundamental class is singleton.

The fundamental relations are used for general definitions. Thus, to define the H_v-field the γ^* is used: An H_v-ring $(R,+,\cdot)$ is called *H_v-field* if R/γ^* is a field [5], and in the sequence the general *H_v-vector space* is defined.

Let $(H,\cdot)$, $(H,*)$ be H_v-semigroups defined on the same set H. $(\cdot)$ is called *smaller* than $(*)$, and $(*)$ *greater* than $(\cdot)$, iff there exists an automorphism $f \in Aut(H,*)$ such that $xy \subset f(x*y)$, $\forall x,y \in H$. Then we

write $\cdot \leq^{*}$ and we say that $(H,^{*})$ *contains* $(H,\cdot)$. If $(H,\cdot)$ is a structure then it is called *basic structure* and $(H,^{*})$ is called *H_b-structure*.

Theorem. Greater hyperoperations of the ones which are WASS or COW, are also WASS or COW, respectively.

Remark 2. The weak axioms lead to a great number of hyper-operations and these hyperoperations define hyperstructures which can be now studied in detail and, in any case, they have a substance; hence they can be considered as hyperstructures with interesting properties. These are many hyperoperations which, in the past, were unlikely to be considered because not even one property was valid in them. We can see that the hyperoperations introduced here are associative only in very special cases and before 1990 such hyperoperations could hardly be considered, even though they appeared in the research. Nevertheless, the created theory can now give results and discover new properties of the obtained hyperstructure. Thus, algebraic domains reveal constructions which seem to be chaotic. Even more so, in certain cases, some of these hyperstructures contain well known structures or hyperstructures, see also [11,12].

This remark follows that constructions and hyper-constructions are needed to be enlarged or to become smaller and we can do this:

Definitions 3. Let $(H,\cdot)$ be a hypergroupoid. We say that *remove* $h \in H$, if we consider the restriction of $(\cdot)$ in $H-\{h\}$. We say that $\underline{h} \in H$ *absorbs* $h \in H$ if we replace h by $\underline{h}$. We say that $\underline{h} \in H$ *merges* with $h \in H$, if we take as product of any $x \in H$ by $\underline{h}$, the union of the results of x with both h, $\underline{h}$, and consider h and $\underline{h}$ as one class, with representative $\underline{h}$.

Most of these constructions are needed in the representation theory. Representations of H_v-groups can be considered either by generalized permutations or by H_v-matrices [6]. The representation problem by *H_v*-matrices is the following:

H_v-matrix is a matrix with entries of an H_v-ring. The hyperproduct of H_v-matrices $\mathbf{A}=(a_{ij})$ and $\mathbf{B}=(b_{ij})$, of type $m \times n$ and $n \times r$, respectively, is a set of $m \times r$ H_v-matrices:

$$\mathbf{A}\cdot\mathbf{B} \;=\; (a_{ij})\cdot(b_{ij}) \;=\; \{\, \mathbf{C}=(c_{ij}) \;\big|\; c_{ij} \in \oplus\Sigma a_{ik}\cdot b_{kj} \,\},$$

where $\oplus$ denotes the *n-ary circle hyperoperation* on the hyperaddition.

Definition 4. Let $(H,\cdot)$ be H_v-group, take a H_v-ring $(R,+,\cdot)$ and a set $\mathbf{M}_R = \{\, (a_{ij}) \,|\, a_{ij}\in R \,\}$, then any map

$$\mathbf{T}: H\to\mathbf{M}_R: h\to\mathbf{T}(h) \;\text{ with }\; \mathbf{T}(h_1h_2)\cap\mathbf{T}(h_1)\mathbf{T}(h_2)\neq\varnothing,\ \forall h_1,h_2\in H,$$

is a *H_v-matrix representation*. If $\mathbf{T}(h_1h_2)\subset\mathbf{T}(h_1)\mathbf{T}(h_2)$, then $\mathbf{T}$ is an *inclusion*, if $\mathbf{T}(h_1h_2)=\mathbf{T}(h_1)\mathbf{T}(h_2)$, then $\mathbf{T}$ is a *good* and an induced representation for the hypergroup algebra is obtained.

In the same attitude recently we defined, using hyperstructure theory, hyperoperations on any type of matrices:

Definition 5 [12]. Let $A=(a_{ij})\in\mathbf{M}_{m\times n}$ be matrix and $s,t\in N$ with $1\le s\le m$, $1\le t\le n$. Then *helix-projection* is a map $\underline{st}:\mathbf{M}_{m\times n}\to\mathbf{M}_{s\times t}: A\to A\underline{st}=(\underline{a}_{ij})$, where $A\underline{st}$ has entries

$$\underline{a}_{ij} = \{\, a_{i+\kappa s,j+\lambda t} \;\big|\; 1\le i\le s,\ 1\le j\le t \ \text{ and } \ \kappa,\lambda\in N,\ i+\kappa s\le m,\ j+\lambda t\le n \,\}$$

Let $A=(a_{ij})\in\mathbf{M}_{m\times n}$, $B=(b_{ij})\in\mathbf{M}_{u\times v}$ be matrices, $s=\min(m,u)$, $t=\min(n,v)$. We define a hyper-addition, called *helix-addition*, as follows

$$\oplus:\mathbf{M}_{m\times n}\times\mathbf{M}_{u\times v}\to\boldsymbol{P}(\mathbf{M}_{s\times t}):(A,B)\to A\oplus B=A\underline{st}+B\underline{st}=(\underline{a}_{ij})+(\underline{b}_{ij})\subset\mathbf{M}_{s\times t}$$

where $(\underline{a}_{ij})+(\underline{b}_{ij})=\{(c_{ij})=(a_{ij}+b_{ij}) \,|\, a_{ij}\in\underline{a}_{ij} \ \text{ and } \ b_{ij}\in\underline{b}_{ij})\}$.

Let $A=(a_{ij})\in\mathbf{M}_{m\times n}$ and $B=(b_{ij})\in\mathbf{M}_{u\times v}$ be matrices and $s=\min(n,u)$. We define a hyper-multiplication, called *helix-multiplication*, as follows

$$\otimes:\mathbf{M}_{m\times n}\times\mathbf{M}_{u\times v}\to\boldsymbol{P}(\mathbf{M}_{m\times v}):(A,B)\to A\otimes B=A\underline{ms}\cdot B\underline{sv}=(\underline{a}_{ij})\cdot(\underline{b}_{ij})\subset\mathbf{M}_{m\times v}$$

where $(\underline{a}_{ij})\cdot(\underline{b}_{ij})=\{(c_{ij})=(\sum a_{it}b_{tj}) \,|\, a_{ij}\in\underline{a}_{ij} \ \text{ and } \ b_{ij}\in\underline{b}_{ij})\}$.

The helix-addition is commutative, is WASS, not associative. The helix-multiplication is WASS, not associative and it is not distributive, not even weak, to the helix-addition. If all used matrices are of the same type, then the inclusion distributivity, is valid.

From the definition of representations by H_v-matrices, we have two difficulties. The first one is to find an appropriate H_v-ring and the

second one is to find an appropriate set of H_v-matrices. However, with the above hyper-multiplication we can use subsets of matrices of type $\mathbf{M}_{m\times n}$ with $m\neq n$. Thus, the representation problem is reduced, as in the classical theory, in searching appropriate sets from usual matrices. This is so, because we have now a hyperalgebra over non-square matrices.

2. New hyperoperations

We will define a hyperoperation in a groupoid equipped with a map f on it. The map plays crucial role so the hyperoperation is called *map* and it is denoted by ∂_f, because the motivation to obtain this is the property which the 'derivative' has on the product of functions. However, since there is no confusion, we will write simply *theta* ∂.

Definition 6. Let $(G,\cdot)$ be a groupoid (respectively, hypergroupoid) and $f:G\to G$ be any map. We define a hyperoperation (∂), we call it *theta-operation*, on G as follows

$$x\partial y = \{f(x)\cdot y, x\cdot f(y)\} \text{ (respectively, } x\partial y = (f(x)\cdot y)\cup(x\cdot f(y))$$

If $(\cdot)$ is commutative then (∂) is also commutative. If $(\cdot)$ is a COW hyperoperation, then (∂) is also COW hyperoperation.

Remark. One can use instead of single valued map f, a multivalued map as well. We will not consider this problem here.

Remark. Motivation for this definition was the map 'derivative' where only the multiplication of functions can be used. In other words, if we 'do not know' the addition of functions. Therefore, for any functions s(x), t(x), we have $s\partial t=\{s't, st'\}$ where $(')$ denotes the derivative.

Properties 7. If $(G,\cdot)$ is a semigroup then:

(a) For every f, the hyperoperation (∂) is WASS.

(b) If f is homomorphism then, again, (∂) is WASS.

(c) If f is homomorphism and projection, or idempotent, i.e. $f^2=f$, then (∂) is associative.

Proof.

(a) For all x,y,z in G we have

$(x\partial y)\partial z = \{f(x)\cdot y, x\cdot f(y)\}\partial z =$

$= \{ f(f(x)\cdot y)\cdot z, (f(x)\cdot y)\cdot f(z), f(x\cdot f(y))\cdot z, (x\cdot f(y))\cdot f(z) \} =$

$= \{ f(f(x)\cdot y)\cdot z, f(x)\cdot y\cdot f(z), f(x\cdot f(y))\cdot z, x\cdot f(y)\cdot f(z) \}$

$x\partial(y\partial z) = x\partial\{f(y)\cdot z, y\cdot f(z)\} =$

$= \{ f(x)\cdot(f(y)\cdot z), x\cdot f(f(y)\cdot z), f(x)\cdot(y\cdot f(z)), x\cdot f(y\cdot f(z)) \} =$

$= \{ f(x)\cdot f(y)\cdot z, x\cdot f(f(y)\cdot z), f(x)\cdot y\cdot f(z), x\cdot f(y\cdot f(z)) \}$

Therefore $(x\partial y)\partial z \cap x\partial(y\partial z) = \{ f(x)\cdot y\cdot f(z) \} \neq \varnothing$, so (∂) is WASS.

(b) If f is homomorphism then we obtain

$(x\partial y)\partial z = \{ f(f(x))\cdot f(y)\cdot z, f(x)\cdot y\cdot f(z), f(x)\cdot f(f(y))\cdot z, x\cdot f(y)\cdot f(z) \}$

$x\partial(y\partial z) = \{ f(x)\cdot f(y)\cdot z, x\cdot f(f(y))\cdot f(z), f(x)\cdot y\cdot f(z), x\cdot f(y)\cdot f(f(z)) \}$

So, again $(x\partial y)\partial z \cap x\partial(y\partial z) = \{f(x)\cdot y\cdot f(z)\} \neq \varnothing$ and (∂) is WASS.

(c) If f is homomorphism and projection then we have

$(x\partial y)\partial z = \{ f(x)\cdot f(y)\cdot z, f(x)\cdot y\cdot f(z), x\cdot f(y)\cdot f(z) \} = x\partial(y\partial z)$.

Therefore, (∂) is an associative hyperoperation.

Notice that only projection without homomorpthism does not give the associativity. Commutativity does not improve the results.

3. Properties and characteristic elements.

We will discuss now some properties in the general case where $(G,\cdot)$ be a groupoid and $f:G\to G$ be a map.

Properties 8.

Reproductivity. For the reproductivity we must have

$$x\partial G = \bigcup_{g\in G}\{f(x)\cdot g, x\cdot f(g)\} = G \text{ and } G\partial x = \bigcup_{g\in G}\{f(g)\cdot x, g\cdot f(x)\} = G.$$

Thus, if $(\cdot)$ is reproductive then (∂) is also reproductive, because

$$\bigcup_{g\in G}\{f(x)\cdot g\} = G \quad \text{and} \quad \bigcup_{g\in G}\{g\cdot f(x)\} = G.$$

Commutativity. If $(\cdot)$ is commutative then $x\partial y=\{f(x)\cdot y,\ x\cdot f(y)\}=y\partial x$, so (∂) is commutative. If f is into $\boldsymbol{Z}_G=\{z\in G \mid z\cdot g=g\cdot z,\forall g\in G\}$, the centre of G, then (∂) is a commutative hyperoperation. If $(G,\cdot)$ is a COW hypergroupoid then, obviously (∂) is a COW hypergroupoid.

Unit elements. In order to have a right unit element u we must have $x\partial u=\{f(x)\cdot u,\ x\cdot f(u)\}\ni x$. But, the unit must not depend on the f(x), so we must have $f(u)=e$, where e be unit in $(G,\cdot)$ which must be a monoid. The same it is obtained for the left units. Therefore, the elements of the kernel of f, i.e. u for which $f(u)=e$, are the units of (G,∂).

Inverse elements. Let u be a unit in (G,∂), then $(G,\cdot)$ is a monoid with unit e and $f(u)=e$. For given x in order to have an inverse element x' with respect to u, we must have

$$x\partial x'=\{f(x)\cdot x',\ x\cdot f(x')\}\ni u \quad \text{and} \quad x'\partial x=\{f(x')\cdot x,\ x'\cdot f(x)\}\ni u.$$

So the only cases, which do not depend on the image $f(x')$, are

$$x'=(f(x))^{-1}u \quad \text{and} \quad x'=u(f(x))^{-1}$$

the right and left inverses, respectively. We have two-sided inverses iff $f(x)u = uf(x)$. For example, if u belongs to the centre of G. In some cases, some elements may have a second inverse.

Proposition 9. Let $(G,\cdot)$ be a group and $f(x)=a$, a constant map on G. Then $(G,\partial)/\beta^*$ is a singleton.

Proof. For all x in G we can take the hyperproduct of the elements, a^{-1} and $a^{-1}x$

$$a^{-1}\partial(a^{-1}\cdot x)=\{f(a^{-1})\cdot a^{-1}\cdot x,\ a^{-1}\cdot f(a^{-1}\cdot x)\}=\{x,a\}.$$

thus $x\beta a$, $\forall x\in G$, so $\beta^*(x)=\beta^*(a)$ and $(G,\partial)/\beta^*$ is singleton. q.e.d.

Remark. If $(G,\cdot)$ be a group and $f(x)=e$, then we obtain $x\partial y=\{x,y\}$ which is the smallest incidence hyperoperation.

Remark. Every $f:G\to G$ defines a partition of G by setting two elements x,y in the same class iff $f(x)=f(y)$, we shall call this partition f-partition and we will denote the class of x by $f[x]$. So, in the above Proposition, we have $f[x] = G = \beta^*(x)$ for all x in G.

Proposition 10. Let $(G,\cdot)$ be a group, e the unit, and f homomorphism, then for (G,∂), we have $x\beta f(x)$.

Proof. Indeed $e\partial x = \{f(e)\cdot x, e\cdot f(x)\} = \{x, f(x)\}$. q.e.d.

Obviously we have $x\ \beta\ f(x)\ \beta\ f(f(x))\ \beta...$

Theorem 11. Let $(G,\cdot)$ be a group and f be an homomorphism, then

$$f[x] \subset \beta^*(x) \text{ for all x in G.}$$

Proof. Let $y \in f[x]$, then $f(y)=f(x)$ but from Proposition10, we have

$$x\beta f(x) = f(y)\beta y, \quad \text{so} \quad x\beta^* y. \quad \text{q.e.d.}$$

4. Special cases and applications

In this paragraph we present some applications and we give some examples in order to see that a large field of research is open.

Application 12. Taking the application on the derivative, consider all polynomials of first degree $g_i(x) = a_i x+b_i$. We have

$$g_1 \partial g_2 = \{a_1a_2x+ a_1b_1, a_1a_2x+b_1b_2\},$$

so it is a hyperoperation inside the set of first degree polynomials. Moreover all polynomials $x+c$, where c be a constant, are units.

Application 13. If R^+ be the set of positive reals and $a \in R^+$, then we take the exponential map $x \to x^a$. The theta-operation takes the form $x\partial y = \{ x^a y, x y^a\}$ for all x,y in R^+. The only one unit is the 1. In order to find the inverses x', of the element $x \in R^+$, we must have $x\partial x' = \{x^a x', x(x')^a\} \ni 1$. From which we obtain that for every element x, there are two inverses, the x^{-a} and $x^{-1/a}$.

Example 14. In the group $(\boldsymbol{Z}_5\text{-}\{\underline{0}\},\cdot)$ we consider the map f: $\underline{1}\to\underline{1}$, $\underline{2}\to\underline{2}$, $\underline{3}\to\underline{3}$, $\underline{4}\to\underline{2}$. Then we obtain the multiplicative table

∂	$\underline{1}$	$\underline{2}$	$\underline{3}$	$\underline{4}$
$\underline{1}$	$\underline{1}$	$\underline{2}$	$\underline{3}$	$\{\underline{4}, \underline{2}\}$

$\underline{2}$	$\underline{2}$	$\underline{4}$	$\underline{1}$	$\{\underline{3}, \underline{4}\}$
$\underline{3}$	$\underline{3}$	$\underline{1}$	$\underline{4}$	$\{\underline{2}, \underline{1}\}$
$\underline{4}$	$\{\underline{4}, \underline{2}\}$	$\{\underline{3}, \underline{4}\}$	$\{\underline{2}, \underline{1}\}$	$\underline{3}$

We remark that there exists only one fundamental class. The map-hyperoperation is not associative but it is WASS, because, for example, $\underline{2}\partial(\underline{4}\partial\underline{4}) = \{\underline{1}\}$ and $(\underline{2}\partial\underline{4})\partial\underline{4} = \{\underline{1}, \underline{2}, \underline{3}\}$.

Example 15. Consider the group $(\boldsymbol{Z}_6,+)$ and the map $f:\boldsymbol{Z}_6\rightarrow\boldsymbol{Z}_6:x\rightarrow x^{-1}$. Then the map-operation is given from the table

∂	$\underline{0}$	$\underline{1}$	$\underline{2}$	$\underline{3}$	$\underline{4}$	$\underline{5}$
$\underline{0}$	$\underline{0}$	$\{\underline{1}, \underline{5}\}$	$\{\underline{2}, \underline{4}\}$	$\underline{3}$	$\{\underline{2}, \underline{4}\}$	$\{\underline{1}, \underline{5}\}$
$\underline{1}$	$\{\underline{1}, \underline{5}\}$	$\underline{0}$	$\{\underline{1}, \underline{5}\}$	$\{\underline{2}, \underline{4}\}$	$\underline{3}$	$\{\underline{2}, \underline{4}\}$
$\underline{2}$	$\{\underline{2}, \underline{4}\}$	$\{\underline{1}, \underline{5}\}$	$\underline{0}$	$\{\underline{1}, \underline{5}\}$	$\{\underline{2}, \underline{4}\}$	$\underline{3}$
$\underline{3}$	$\underline{3}$	$\{\underline{2}, \underline{4}\}$	$\{\underline{1}, \underline{5}\}$	$\underline{0}$	$\{\underline{1}, \underline{5}\}$	$\{\underline{2}, \underline{4}\}$
$\underline{4}$	$\{\underline{2}, \underline{4}\}$	$\underline{3}$	$\{\underline{2}, \underline{4}\}$	$\{\underline{1}, \underline{5}\}$	$\underline{0}$	$\{\underline{1}, \underline{5}\}$
$\underline{5}$	$\{\underline{1}, \underline{5}\}$	$\{\underline{2}, \underline{4}\}$	$\underline{3}$	$\{\underline{2}, \underline{4}\}$	$\{\underline{1}, \underline{5}\}$	$\underline{0}$

This is a commutative hyperoperation, it is WASS, because, for example, $\underline{1}\partial(\underline{1}\partial\underline{2}) = \{\underline{2}, \underline{4}\}$ and $(\underline{1}\partial\underline{1})\partial\underline{2} = \{\underline{0}, \underline{2}, \underline{4}\}$, so $(\boldsymbol{Z}_6,\partial)$ is a commutative H_v-group. One can obtain that

$$(\boldsymbol{Z}_6,\partial)/\beta^* = \{\{\underline{0}, \underline{2}, \underline{4}\},\{\underline{1}, \underline{3}, \underline{5}\}\} \cong \boldsymbol{Z}_2.$$

This is not cyclic since $x\partial x = \{\underline{0}\}$ for all x in $\boldsymbol{Z}_6$, i.e. every element has itself as the only one inverse element.

Example 16. Consider the group $(\boldsymbol{Z}_6,+)$ and the map

$$f: \underline{0}\rightarrow\underline{0}, \underline{1}\rightarrow\underline{1}, \underline{2}\rightarrow\underline{2}, \underline{3}\rightarrow\underline{3}, \underline{4}\rightarrow\underline{4}, \underline{5}\rightarrow\underline{2}.$$

Then the map-operation is given from the table

∂	$\underline{0}$	$\underline{1}$	$\underline{2}$	$\underline{3}$	$\underline{4}$	$\underline{5}$
$\underline{0}$	$\underline{0}$	$\underline{1}$	$\underline{2}$	$\underline{3}$	$\underline{4}$	$\{\underline{2}, \underline{5}\}$
$\underline{1}$	$\underline{1}$	$\underline{2}$	$\underline{3}$	$\underline{4}$	$\underline{5}$	$\{\underline{0}, \underline{3}\}$
$\underline{2}$	$\underline{2}$	$\underline{3}$	$\underline{4}$	$\underline{5}$	$\underline{0}$	$\{\underline{1}, \underline{4}\}$
$\underline{3}$	$\underline{3}$	$\underline{4}$	$\underline{5}$	$\underline{0}$	$\underline{1}$	$\{\underline{2}, \underline{5}\}$
$\underline{4}$	$\underline{4}$	$\underline{5}$	$\underline{0}$	$\underline{1}$	$\underline{2}$	$\{\underline{0}, \underline{3}\}$
$\underline{5}$	$\{\underline{2}, \underline{5}\}$	$\{\underline{0}, \underline{3}\}$	$\{\underline{1}, \underline{4}\}$	$\{\underline{2}, \underline{5}\}$	$\{\underline{0}, \underline{3}\}$	$\underline{1}$

One can obtain that

$$(\boldsymbol{Z}_6,\partial)/\beta^* = \{\{\underline{0}, \underline{3}\},\{\underline{1}, \underline{4}\},\{\underline{2}, \underline{5}\}\} \cong \boldsymbol{Z}_3.$$

$(\boldsymbol{Z}_6,\partial)$ is a cyclic H_v-group where $\underline{1}$ and $\underline{5}$ are generators of period 5.

Example 17. Consider the group $(\boldsymbol{Z}_6,+)$ and the map

$$f:\ \underline{0}\to\underline{0},\ \underline{1}\to\underline{1},\ \underline{2}\to\underline{2},\ \underline{3}\to\underline{3},\ \underline{4}\to\underline{2},\ \underline{5}\to\underline{5}.$$

Then the map-operation is given from the table

∂	$\underline{0}$	$\underline{1}$	$\underline{2}$	$\underline{3}$	$\underline{4}$	$\underline{5}$
$\underline{0}$	$\underline{0}$	$\underline{1}$	$\underline{2}$	$\underline{3}$	$\{\underline{2}, \underline{4}\}$	$\underline{5}$
$\underline{1}$	$\underline{1}$	$\underline{2}$	$\underline{3}$	$\underline{4}$	$\{\underline{3}, \underline{5}\}$	$\underline{0}$
$\underline{2}$	$\underline{2}$	$\underline{3}$	$\underline{4}$	$\underline{5}$	$\{\underline{0}, \underline{4}\}$	$\underline{1}$
$\underline{3}$	$\underline{3}$	$\underline{4}$	$\underline{5}$	$\underline{0}$	$\{\underline{1}, \underline{5}\}$	$\underline{2}$
$\underline{4}$	$\{\underline{2}, \underline{4}\}$	$\{\underline{3}, \underline{5}\}$	$\{\underline{0}, \underline{4}\}$	$\{\underline{1}, \underline{5}\}$	$\underline{0}$	$\{\underline{1}, \underline{3}\}$
$\underline{5}$	$\underline{5}$	$\underline{0}$	$\underline{1}$	$\underline{2}$	$\{\underline{1}, \underline{3}\}$	$\underline{4}$

One obtains that

$$(\boldsymbol{Z}_6,\partial)/\beta^* = \{\{\underline{0}, \underline{2}, \underline{4}\},\{\underline{1}, \underline{3}, \underline{5}\}\} \cong \boldsymbol{Z}_2.$$

For the reproductivity, the element $\underline{4}+\underline{4}$ which does not appeared in the normal position in the result it appears, in the general case, as follows:

$$x+x\in x\partial(x+x-f(x))=\{f(x)+x+x-f(x),\ x+f(x+x-f(x))\},\ \forall x\in \boldsymbol{Z}_6,$$

so the reproductivity is clear.

We conclude with a theorem on this field.

Theorem 18. Consider the commutative group of integers $(\boldsymbol{Z},+)$ and let $n\neq 0$ be a natural number. Take the map f such that $f(n)=0$ and $f(x)=x$ for all x in $\boldsymbol{Z}-\{n\}$. Then

$$(\boldsymbol{Z},\partial)/\beta^* \cong \boldsymbol{Z}_n.$$

Proof. First, for all x,y in $\boldsymbol{Z}-\{n\}$ we have, for the theta-operation,

$$x\partial y = \{f(x)+y,\ x+f(y)\} = \{x+y\},$$

so the hypersum is a singleton and coincides with the usual sum in $\boldsymbol{Z}$.

For all x in $\boldsymbol{Z}-\{n\}$ we have

$$x\partial n = n\partial x = \{f(x)+n,\ x+f(n)\} = \{x+n,\ x\}.$$

Finally $n\partial n = \{f(n)+n,\ n+f(n)\} = \{n\}$.

Therefore $x\beta(x+n)$. Moreover, from the above, we obtain that for all x,y in $\boldsymbol{Z}$, the hypersum $\{x,\ x+n\}\partial\{y,\ y+n\}$ belongs to the same class modn. Thus, the fundamental classes are the classes modn.

Therefore $(\boldsymbol{Z},\partial)/\beta^* \cong \boldsymbol{Z}_n$. q.e.d.

Remark that this construction is an analogous case to the case of the uniting the elements 0 and n, see [6].

References

[1] Corsini, P., *Prolegomena of Hypergroup Theory,* Aviani Editore, 1993.
[2] Corsini, P., Leoreanu, V., *Applications of Hypergroup Theory,* Kluwer Academic Publishers, 2003.
[3] Davvaz, B., *On H_v-rings and Fuzzy H_v-ideals*, J.Fuzzy Math. V.6, N.1(1998),33-42
[4] ___, *Abrief survey of the theory of H_v-structures,* 8th AHA Congress, Spanides Press (2003),39-70.
[5] Vougiouklis, T., *The fundamental relation in hyperrings. The general hyperfield,* Proc. 4thAHA, World Scientific (1991),203-211.
[6] ___, *Hyperstructures and their Representations,* Monographs in Mathematics, Hadronic Press, 1994.
[7] ___, *Some remarks on hyperstructures,* Contemporary Mathematics, Amer. Math. Society, 184,(1995), 427-431.
[8] ___, *Constructions of H_v-structures with desired fundamental structures,* New frontiers in Hyperstructures, Hadronic Press (1996), 177-188.
[9] ___, *Enlarging H_v-structures,* Algebras and Combinatorics, ICAC'97, Hong Kong, Springer - Verlag (1999),455-463.
[10] ___, *On H_v-rings and H_v-representations,* Discrete Mathematics, Elsevier, 208/209 (1999), 615-620.
[11] ___, *Finite H_v-structures and their representations,* Rendiconti Seminario Matematico di Messina S.II, V.9 (2003),245-265.
[12] Vougiouklis, T., Vougiouklis, S., *The helix hyperoperations,* to appear in Italian Journal of Pure and Applied Mathematics.

Euclid and the scientific thought in the third century B.C.[1]

Renato Migliorato
Giuseppe Gentile[2]
Department of Mathematics, University of Messina
Address: Contrada Papardo, 98121 Messina (Italy).
E-mail: renato.migliorato@unime.it, gentile@dipmat.unime.it

Abstract

The criticism on the texts of Euclid, even assuming different positions, starts generally from the previous assumption that the author of the Elements is totally inside the Platonic-Aristotelian tradition. The thesis affirmed in this paper is that many of the gaps and contradictions found by the criticism have their root in this assumption. The authors assert that Euclid was a scientist that belonged in a full way to the new cultural climate of the Hellenistic Kingdoms, and particularly of the Alexandria's Museum. In this climate, characterized by lively philosophical disputes, the scientists, and in particular Euclid, tend to obtain coherent and stable results, voluntarily omitting to give their opinion on the real being of the scientific object and on the truth of the principles.

1. Introduction

Even if important innovations in the critical studies on Euclidean geometry don't less in a more recent times, the period starting from the end of XIX century until the beginning of the XX century is surely the more prolific one, that in which a critical order was constituted such to be considered until now *almost* definitive. From 1850 to 1928 the Heiberg and Menge's edition[3] of the Euclid's works was published; this is considered the more reliable text and the nearest to the original one. In the first half of the twentieth century there are many translations, with comments and remarks, founded on the Heiberg's text, as that one of Federigo Enriques[4] or that one of Heath[5], and many critical elaborations by the same and other authors[6]. Among the more recent published works that propose some new interpretative hypotheses, it seems suitable to us to men-

[1] Work supported by University of Messina as Local Research Project (P.R.A.).
[2] This work, coordinated by R. Migliorato, is the result of a collaboration that sow contemporarily engaged both the authors on all the treated aspects. Also the searches that separately were effected, was discussed in all points before to decide which solution to adopt. The contribution of G. Gentile, concretised however prevalently on the sections 2, 3 and 5, can be quantify as a third of the whole work.
[3] EUCLIDES, 1883-1916.
[4] ENRIQUES, 1912-1935.
[5] HEATH, 1956.
[6] See for instance TANNERY; VAILATI; VERONESE; AMALDI; ENRIQUES, 1912.

tion two articles and a monograph of Lucio Russo[7], a monograph of Francesca Incardona[8] and a book of Imre Toth[9]. The articles of Russo consider the first seven definitions in the first book of the *Elements*, that he affirms to be posterior interpolations. With that, it would come to fall the greatest residual reasons in favour of a supposed Platonism of Euclid. We believe that such hypothesis, well argued and documented by the author, must be accepted, not only because it is historically reliable, but also because it seems to us most suitable to answer to difficult problems, remained unsolved, on the interpretation of the Euclidean work.
The book "*La rivoluzione dimenticata*" (*The forgotten revolution*), also by Russo, would demand instead a more complex and articulated valuation that, in its entirety, is extraneous to the object and the purpose of the present paper. However the fundamental hypotheses that are the nucleus of the book, cannot remain excluded from our analysis.
For our purposes, it seems meaningful to us the introductory text by which Incardona accompanies her translation of Euclid's *Optic*. It seems interesting to us, in particular, the reasonings that leads to interpret the work as a mathematical model of the phenomenon of the vision and tend to insert Euclid in the context of a more wide evolution of the scientific and philosophical ideas at the beginning of third century. Not well founded it appears instead the hypothesis, however only fleetingly pointed out by the author, that Euclid could be absolutely assimilated to the area of the Stoic philosophy. Finally the book of Imre Toth[10] is interesting for us because it calls our attention on some passages of Aristotle that evidence the existence, already before Euclid, of an open problem on the parallel straight lines.
The purpose of the present paper is a re-examination of the Euclidean text to the light of the last criticism's history, with particular reference to the mentioned works, and of the most recent acquisitions of knowledge on the Hellenistic society and culture of third century B.C. The conclusions to which we will reach, seem to answer to some questions that still now remained opened.

2. Platonist or Aristotelian?

The date of the *Elements* is set about 300 B.C. and a more precise dating is objectively difficult. But if we look at the indications of Proclus[11], still now the main source about this argument, we have to think that he written his works in the first decades of the third century. In effect, the contemporaneousness with the First Ptolemy imposes only that Euclid was present and active before 283 B.C., date of the death of Ptolemy, but this doesn't tell something of more precise about the period during which his work was elaborated; so, as principle, a lot of hypotheses would be possible. Everything however allows us to suppose that the Elements and the other principal known works of Euclid were written within the Museum and the Library of Alexandria.

---

[7] RUSSO, 1992, 1997, 1998.
[8] INCARDONA, 1996..
[9] TOTH, 1997.
[10] It is difficult to summarize in a little space the content of an ample monograph that tends to a re-interpretation of the Aristotelian thought to the light of the not-Euclidean geometries.
[11] *".....Not much younger than these* [Hermotimus of Colophon and Philippus of Medma] *is Euclid, who put together the Elements. collecting many of Eudoxus theorems, perfecting many of Theaetetus, and also bringing to irrefragable proof the things which were only somewhat loosely proved by his predecessors. This man lived in the time of the first Ptolemy. For Archimedes, who lived immediately after the first (Ptolemy), makes mention of Euclid: and further, they say that Ptolemy once asked him if there was in geometry any shorter way than that of the elements, and he answered that there was no royal road to geometry. He is then younger than the pupils of Plato but older than Eratosthenes and Archimedes; for the latter were contemporary with one another as Eratosthenes somewhere says. For his ideas Euclid was platonic and had very families this philosophy, so much that the final purpose of the whole collection of Elements was the construction of the so-called platonic figures."* (PROCLUS, Comm. Eucl.., II, 68).

For vastness and organization, but also for its intrinsic characters that we will see more, such work seems fully to reconcile with a great organized enterprise, what the school in Alexandria was certainly, and with a historical phase in which the sciences were detached by the philosophy and this also has the tendency to abandon the great systems to follow more pragmatistic and empiristic ways[12].

Regarding the supposed Platonism of Euclid, affirmed with so much safety by Proclus, we observe that this is a partial and unreliable interpreter in order to this problem: his marked new-Platonic positions can have induced him to suppose, and then also to find, evidences of Platonic thought in a scientist that was certainly considered as the greatest mathematician of every time[13]. Of course it was difficult to find contrary evidences, just for the lack of explicit affirmations and comments that could reveal his philosophical options.

Someone, in the XIX century affirmed that "*For Proclus Euclid had the great fortune not to be denied neither from the Caldean oracles, neither from the speculations of the old and new Pythagorean philosophers* "[14]. But as we will say forward, we can well hypothesize that the apparent aseptic style of the Euclidean writings and other scientific works of the same century, are not determined by the fortune but by a precise choice; and it is possible that such hypothesis furnishes a key of reading of the Alexandrine science, and at the same time an explanation of the survival of several scientific texts, while all the original texts of the same period that subtend a vision of the world are lost[15].

Naturally what we said, doesn't confirm but doesn't even exclude an adhesion or a proximity of Euclid to the Platonism. And so the modern criticism, even if it doesn't consider seriously the

---

[12] For instance Ludovico Geymonat wrote "*The museum of Alexandria represents the triumph of the specialized culture: the so-called Hellenistic culture. The field of the to knowledge is divided in well circumscribed departments. Neither general philosophical systems nor vast syntheses are created, but rigorous researches are developed on particular problems facing them one for time. The whole type of teaching has the tendency just to form some researchers more rich in serious and sure doctrine. This tendency towards the detail explains the interest for the scientific investigation and, still more, the method of the specialization adopted by the Hellenistic science. While the great philosophers treated, with equal boldness and competence, of physics and of mathematics, as Plato did, for example, or of logic and natural sciences as Aristotle and Theophrastus, the scientists of the Hellenistic age are not expert in philosophy, and inversely the philosophers neglect scientific investigation, to reduce himself to his own specific competences*" (GEYMONAT, p. 284). And Enriques: "*Later, with the diffusion of the Greek civilization that was subsequent to the Macedonian conquests, other centres, splendid of culture, flourished, as Rhodes, Pergamos, and above all Alexandria of Egypt. In this Hellenistic period the science is loose from the philosophy (that it seems by now exclusively dominated by moral interests) and it touches to a florid maturity: to which nevertheless the decadence follows soon, also if slowly*". (ENRIQUES 1925, 1, p. 15).

[13] The affirmation that the purpose of the Elements is constituted by the platonic bodies (the five regular polyhedrons), doesn't have however some bases, because although they constitute the last matter of the work, there are remarkable parts of this that haven't any application to the regular polyhedrons. The same affirmation becomes laughable if we think that the principles and the theorems of the Elements are used in following works, particularly in the *Optic* and this, together with the notions of the *Elements*, is applied to the *Phenomena*. If we use the criterion of Proclus then we should conclude instead that the *Elements* have as finality to "*save the phenomena*". (See more forward, the section 4).

[14] "*Martin says rather neatly, «Pour Proclus , les Éléments d'Euclide ont l'heureuse chance de n'être contredits ni par les Oracles chaldaïques, ni par les spéculations des pythagoriciens anciens et nouveaux...* »." HEATH, 1956, I, p. 30, note 2.

[15] The authors prefer, for the time being, to suspend the judgment on the reasons and the dynamics that will bring, with the crisis and the following end of the Hellenistic kingdoms, to a deep inversion of tendency in the vision of the world; but it is enough evident that in the immediately following centuries the philosophies of this period are strongly opposed. It is enough to think that all testimonies on the ancient Stoa (from Zeno of Citium to Chrysippus of Soli) are in antagonistic key and often (see as ex. Galen, but also Cicero) denounce an impassioned aversion or quite acridity. But not less impassioned criticism appears towards all that texts (not only philosophical) of the third century by which a conception of the world or a theory of the knowledge could be deduced.

affirmation of Proclus[16], it doesn't always exclude an ascendancy of Platonic origin on the Euclidean geometry, but rather it sometimes uses categories proper of the founder of the Academy in the analysis of the Elements. Heat said on this argument:

> "*He may himself have been a Platonist, but this does not follow from the statements of Proclus on the subject* […] *It is evident that it was only an idea of Proclus own to infer that Euclid was a Platonist because his Elements end with the investigation of the five regular solids, since a later passage shows him hard put to it to reconcile the view that the construction of the five regular solids was the end and aim of the Elements with the obvious fact that they were intended to supply a foundation for the study of geometry in general*"[17].

Enriques observes instead that

> "*Plato* (Resp. *527) seems to disdain the use of postulates, where he criticizes the «too much ridiculous and miserable terminology» of the geometers, which, as it was treated of practical purpose, they speak always to square, to prolong, or to add, while the whole science is cultivated to the purpose to know"* [18].

Effectively, from a Platonist man we could expect at least a different formulation of the first three postulates that clearly introduce in the geometry a constructive character! Neugebauer is instead very drastic; he denies in general whatever influence of Plato on the mathematicians:

> *"It seems evident to me that the role of Plato has been enormously exaggerated. His direct contributions to the mathematical knowledge have openly been void. The fact that, for brief time, mathematicians of the level of Eudoxus belong to his entourage, is not a proof of the influence of Plato on the mathematical research.* […] *The doctrines of Plato have, without doubt, practiced a great influence* ***on the modern interpretation*** *of the Greek science. But if the modern studious had devoted to Galen or to Ptolemy as more attention as to Plato and his followers, they would rather have reached different results and would not have invented the myth of the strong attitude of the so-called Greek spirit to develop scientific theories without making petition to experiments or to empiric verifications"*[19].

A point remains to examine, that in our opinion is very important, it is that concerning the first seven definitions or terms (ὅροι). These, according to someone, would be able to confirm the Platonist thesis. But their uselessness is noticed by several commentators; in fact they are never used in the proofs of theorems. Moreover, not always they are clear, as the terms 4 and 7, in contrast with the extreme linearity and transparency of the work in every other part[20]. The attempts to explain such definitions, and particularly the 4 and the 7[21], already mentioned, afar from to do clarity, create more problems than they solve. On this matter an answer can come from the already mentioned paper of Russo[22] that by an ample analysis of the texts and the sources, reaches the conclusion that the ὅροι 1-7 would probably be formulated by Heron and, however, added in a later period, presumably for educational or explanatory motives. Beyond the convincing historical reasonings, this hypothesis allows us to answer to questions that otherwise would be dark. So we are inclined for the acceptance of this thesis, but we have to say that the matter cannot be

---

[16] Nevertheless exposures don't miss, mostly of synthetic, encyclopaedic or educational character, that in absence of further precise statements, can confirm the thesis of Proclo or an anterior dating to the 300 B.C. We mention only an example in electronic edition and therefore recent. To the item Euclid (Marinus Taisbak, in ENCYCLOPÆDIA BRITANNICA, on line edition, Britannica.com Inc., 1999-2001) we can read: "*For his subject matter Euclid doubtless drew upon all his predecessors, but it is clear that the whole design of his work was his own, culminating in the construction of the five regular solids, now known as the Platonic solids*."

[17] HEATH, 1956, vol. I, p. 2.

[18] ENRIQUES, , 1912, I, p. 42.

[19] NEUGEBAUER, pp. 183-184.

[20] An exception is perhaps constituted by the following def. 8 and 9 and some others. But on this topic we will speak forward.

[21] See for ex.. ENRIQUES, 1912, I, p. 30 ; HEATH, 1956, pp. 155-176.

[22] RUSSO, 1998.

considered as definitely closed; in fact, if Euclid has left indefinite the fundamental geometric notions, why would had to define the concepts of unity and that one of number (ὅροι 1 and 2 of Book VII)? This introduces a possibility of a more general discourse extended to these and perhaps other definitions. However there are strong signs of possible additions or changes during the time: for example the definition of the *odd time odd numbers*[23], not only is useless for itself, but it doesn't appear in the Teon's edition[24].

We want only finally discuss the problem, already mentioned[25], of the ὅροι 8 and 9, on which numerous questions have also risen[26]. In first position it is the reason for which Euclid owes to give a general definition of angle, containing also that one of curved angles. It is true that before Euclid such angles was considered, but in the *Elements* the more general definition remains an aim for itself, because, as definitions 1-7, they are never used. We can hypothesize that also the ὅροι 8 and 9, at least in the form in which they appear in the text of Heiberg, has been written in a posterior period. In this case however we can ask to us if the angle was originally treated as not defined term or if Euclid fixed one only definition (and what) for the rectilinear angle. The actual definition, moreover, doesn't define anything because it is limited to set a relationship of synonymy between the words *angle* and *inclination*. It is not possible, in this paper, to give an answer to a question that would require a specific research but in every case it deserves to be deepened.

At this point we can affirm, without further hesitations, that anything is found in the Elements that can confirm a direct descent of the Euclid's thought from that one of Plato, neither a such influence can be historically documented. There are contrarily a lot of reasons to think Euclid enough away from a Platonist conception of the science.

More complex is the matter of the relationship between Euclid and Aristotle. While, in fact, many aspects of the Aristotelian theory of the knowledge are present, in different form and measure, in the whole history of the science, from Euclid to our days[27], on the other hand it is difficult today to understand what Euclid knew of the Aristotelian doctrine that is also known to us[28]. It is permissible nevertheless to suppose that the Aristotelian thought, anyway, had its diffusion and a role of primary importance in the formation of the school of Alexandria, also for the

---

[23] EUCLIDES, Elem. VII, *def.* 10.

[24] See. ENRIQUES, 1912-1935, II, p. 170.

[25] V. nota 20.

[26] V. ad es. ENRIQUES, 1932, I, pp. 32-35 e HEATH, 1956, I, pp. 176-178.

[27] It is true that the birth of the modern science is considered as the fruit of the Galileo's opposition to the Aristotelian tradition, but it is also true that in every phase and in every passage of its development, the scientific thought had to confront itself with the great themes set by Aristotle. Among these themes we remember particularly the distinction between philosophy and particular sciences and the definition of scientific knowledge that ever since is indissolubly tied up to the concept of cause. For Aristotle to know (in a not sophistic way) means in fact to know "the cause by which a thing **is** and **can't not to be**" (Post. Anal. I, 2; Metaph, I, 1). The simple description of things or facts or the presence of contingent relations doesn't re-enter therefore in the scientific knowledge. The search of the general causes for a class of phenomena was the base of all the development of the ancient (post-Aristotelian) and modern sciences, even if the concept of cause lose, with Hume, his metaphysic mean (David Hume: *A Treatise of Human Nature*). When the final cause was expunged from the modern sciences, only the concept of efficient cause remained and produced the consequent deterministic vision of the nature. It can seem that the modern science, refusing the final cause, keeps, clearly and definitively distance from Aristotle at least on this. But the birth of the cybernetics has forced the attention of the scientists on the "finalistic behaviours", simulated until now by the mechanism of the feed back, and even if this is conceived in reality as a deterministic mechanism that simulates in the results only a finalistic behaviour, different conceptions there are such that can to stake the whole matter (on the subject see for instance H. VON FOERSTER: *La verità è l'invenzione di un bugiardo*, Melteni, Roma, 2001).

[28] Notoriously the so-called esoteric or achromatic papers of Aristotle, and this means almost all the work that we today know, were found again to Athens and subsequently, brought to Rome by Silla, they were ordered and published by Andronicus of Rhodes in the first century B.C. Euclid therefore could not have read, of Aristotle, the same works that we know and on which in the centuries the whole exegesis of the Aristotelian thought was founded; however it is likely that he knew, by other ways, the Aristotelian teaching.

prestige of which the philosopher of Stagirus had to enjoy to the successors of Alexander. So if it is true that *to bring the Aristotelian thought to account* has been an unavoidable necessity of the whole scientific thought, then it is also true that we don't have now to establish if Euclid *was Aristotelian*, but if his work introduced elements of novelty and originality in comparison to the past and particularly in comparison to the philosopher of Stagirus.

To such purpose first of all we will try to individuate what characters of the work of Euclid are surely maintained in the line of the Aristotelian science. These characters can be easily identified with the deductive structure, that even if it was somehow already delineated by precedents geometers, nevertheless Aristotle was he who clearly theorized and systematized it. To this deductive structure the name *Elements* (Στοιχεία) surely alludes[29] .

For Aristotle *elements* are the indivisible components of something to which they are immanent[30], so in the case of the Euclidean Geometry, we can think that the "*elements*" to which the title alludes is (1) Components of the geometric objects, that is solids, surfaces or lines, with exclusion of the points that cannot have components. Or (2) Components of proofs.

The first interpretation is given by different commentators, according to which, points, lines, and surfaces would be the elementary components to which the title would point out. This interpretation however doesn't convince. Already in the Platonic conception, the mathematical beings, because they are ideas, are for itself indivisible. Also in the Aristotelian conception, all the objects of the geometry, except the point, are of course divisible as greatness, but they are not divisible as notion, as definition and as category[31]. So a surface divides a body that has the form of a geometric solid, but it is not a constituent element of this last, neither as notion, neither as definition or category (or as idea for the supporters of Platonic thesis). Equally a point is the limit or the division of a line but it is not an element of it[32]. Besides to consider the points as constitutive elements of a geometric figure would involve the admission of the actual infinite. This is expressly excluded by Aristotle and, thing more important for us, carefully avoided by Euclid. Moreover the fundamental objects as point, line, surface, are not the real object of the Euclid's *Elements*; there the whole plain geometry, the solid one, the theory of numbers and the theory of proportions are developed, while on such fundamental objects too much little is said, or nothing if the first seven ορόι are considered apocryphal.

The second interpretation (always following Aristotle), would concern the first principles of the geometry, that is the postulates and, at the most, the common notions. These in fact are, for Aristotle, the indivisible constituent parties of the proofs.

---

[29] It isn't important if this title was or not assigned by Euclid himself; we are interested to observe that, as Proclus said, before Euclid others geometric works was already called *Elements*, in particular that ones written by Hyppocrates of Chios (470-410).

[30] Aristotle articulates this definition in cases and sub-cases but all belonging to the same general concept that presupposes: (1) to be parts of something, (2) to be indivisible (See Metaph., 1014 a, 26 –1014 b, 15.) .

[31] See. Metaph. 1016b.

[32] Aristotle defines the surface as limit or division of a solid, the line as limit or division of a surface and the point as limit or division of a line (for ex. "*…when the bodies are set to contact or are divided, in the moment in which one touches another, one surface only is formed and, in the moment they are divided, they form two surfaces. Accordingly, when the bodies are gathered, the two surfaces don't exist anymore and they result destroyed; when, instead, the bodies are separated, the two surfaces exist as first they didn't exist* […]. *In fact, all these things* [lines and surfaces] *are, in the same way or limits or divisions"* Metaph, 1002a, 39–102b, 10). These entities are not able however to exist, for Aristotle, separated by the bodies, and only our mind is able to consider it separately and independently from them (for ex.: "*It is shown therefore sufficiently that the mathematical being are not substances in taller degree of the bodies, and that, in comparison to the sensitive ones, they don't have a priority in the order of the notion and, finally, that are not able in some way to separately exist*" Metaph. 1077b, 12-15, and "*So the mathematical sciences won't be sciences of sensitive things, but they won't be even sciences of other objects separated by the sensitive ones*" Ibid. 1078a, 3-5)

As we will see this conception finds some difficulties if we consider possible to assume as postulate indifferently one or another between two equivalent propositions. But as we will say forward, this position doesn't seem to be that one of Aristotle because he supposes instead a fundamental asymmetry between what is proved (more complex) and what needs to prove it (more simple), until to arrive so to a base of not demonstrable propositions: only these would be therefore elements. In the Euclidean Geometry, instead, the fifth postulate, at least, doesn't seem to satisfy to this condition and this set some difficulties for the second interpretation also.

But there is a third hypothesis: that Euclid uses the word Στοιχεία following not Aristotle but a precedent tradition in which *elements* means of course *the parts that constitute a proof* but without the pretension that they are not demonstrable. In this way, the greatest part of theorems, besides the postulates, would be elements of other theorems in whose proofs they are used. This hypothesis could justify the choice of the term as title of a work that is characterized for a branched structure according to a well precise definable relationship of partial order as "*A is element of B*" and the "*postulates"* constitute a set of minimal elements. This hypothesis is better suitable if we also consider the fact that the proof of a theorem was conceived, in the Greek tradition and by Aristotle also, as a decomposition of a proposition in several more simple propositions that one uses in the proof. The fact that already before Euclid some *Elements* have been written, confirms this interpretation, because in the Greek tradition the first proofs should concern the most complex and important propositions, as are those on the greatness and those concerning relationships (equalities, similarity, theorem of Pythagoras, etc…) while the proofs were founded upon propositions more reliable about which there were not doubts. The gradual refinement of the critical analysis should have induced greater caution and smaller confidence towards the presumed truths that were previously admitted, until the constitution of an inductive chain, and so, because an endless chains is impossible, the search had beginning of the simplest "*truths*". We can read on this subject the following passage by Enriques:

> *"…in the work of Euclid…in fact appear to the first places theorems - as those on the equality of the triangles - that aren't able to belong to a primitive period of the geometric development because they haven't meant for itself, but receive it only as begins or elements of a chain that conducts to really meaningful geometric properties: as the sum of the angles of the triangle and the relationship (Pythagorean) among the squares of the sides of the right-angled triangle, the two fires to which the arrangement of the first book aims"*[33].

The deductive order is therefore a sure element of continuity with the Aristotelian thought, but partly also with the tradition of the Greek geometry, at least beginning from Hyppocrates of Chios[34]. And certainly it is not easy to distinguish the two aspects, also because Aristotle himself broadly uses the logical structures already consolidated within the mathematical studies, as bricks to compose his philosophical system. This continuity with the preceding tradition is clearly implicit in the passage of Proclus in which, coherently with our conclusions, the meaning of the word "*Elements*" is explained as follows:

> *"…Besides, the term «element» can be used in two senses, as Menaechmus says: …, what proves is element of what is proved, as in Euclid the first proposition is element of the second one and fourth one of the fifth one…"*[35]

It is interesting the reference to Menaechmus because it proves that surely before Euclid the term was really used with the meaning that we said.

---

[33] ENRIQUES, *L'evoluzione etc.*, 1912, p. 4.

[34] We remember that Proclus said that the use of the term *Elements*, dates back at least to Hyppocrates of Chios (470 - 410 B.C). See PROCLUS, *Comm.* II, IV, 66.

[35] PROCLUS, *Comm.*, II, VII, 72.

3. The Hellenistic science.

The problem regarding the role that the particular sciences had in Hellenistic age, and particularly from the third century to the second one B.C., is still away from to have a satisfactory and univocal solution. It is broadly shared common notion that after the peak of the Alexandrine civilization, and precisely from the second half of the second century B.C., a phase beginnings of decadence, while the range and the extension of the scientific enterprise, besides the moment and the reasons for which this development would be interrupted, appear still problematic. The thesis is very accredited for which the Hellenistic science would have reached the threshold of a scientific revolution, but without never crossing it, although there were (in everything or partly) the theoretical premises. The literature on this matter is very wide; it will be enough to quote some examples only. The first one is quoted by Ludovico Geymonat that wrote as follows:

> "*In front of the first victorious affirmations of such method* [application of the scientific principles to the technology], *that becomes today the main base of the modern technical civilization, there is to wonder for what motive it has not had in the antiquity a greater development, and it remained instead conscripted to some isolated cases* […]. *It is a very complex problem, that in general way can be formulated as follows: Way not even a sketch of mechanical civilization was not developed, in the ancient world, while undoubtedly there were the first theoretical premises, though in limited measure?* […]. *The cause* […] *probably can be found in the social structure of the Greek-Latin world, which didn't feel the need to invent new machines, sufficiently having already to own cheap disposition - the «natural machines» of the slavery.- We remember on such matter that Marcus Terentius Varro, describing the tools by which the earth is worked, he textually reports that «somebody divide them in three categories: speaking tools, semi-speaking-tools and mute tools"*[36].

The second example that we want to quote is by Federigo Enriques and Giorgio de Santillana which affirm:

> "*But who wants to understand the motives for this superb flowering* [the extraordinary cultural climate that was developed around the museum of Alexandria] *is induced to look, over the external environment, the intimate conditions of the work of the researchers: as we said the thought, forgotten the universalistic claims, is now circumscribed within specific fields of search and, on the base of simple postulates, it succeeds in answering to determined problems. This separation from the philosophy seems a liberation of the science that, renouncing to really know the nature of the things, acquires properties of its real object and tries to derive the most important positive results. At the same time the great means of study, the most frequent contacts of the reunited researchers in the Museum and the practice of the teaching that disciplines together teachers and pupils, compete to form a school in the modern sense: not longer philosophical school, that receives impulse from the metaphysical idea of a Chief, but scientific school where different intelligence unite their efforts, creating and preserving the tradition of the method. Nevertheless it is easy to imagine that these reasons can't be enough for long time to maintain the progress of the science, if the interest of the problems is not relighted by an always living philosophical vision, and the work of the narrow class of researchers doesn't feed from an underlying culture of the people. Under such conditions cannot surprise that the flowers quickly budded of the scientific genius come soon to fade."*[37].

---

[36] GEYMONAT, 1973, p. 300. We has to observe as Geymonat uses in this passage two criterions that we think debatable: a generic reference to a Greek-Latin world, without further precise space-temporal references and the appeal to a few pertinent source. Although in fact Terentius Varro is not entirely out of the considered temporal arc, he purely belonged to a cultural Latin and Roman circle, that between II and I centuries B.C. was still set in comparison to the Hellenistic world in terms of conquest and therefore of difference. But on this theme we will return forward.

[37] ENRIQUES, SANTILLANA, 1937, p. 148.

On the same theme, the following passage of Ludwig Edelstein seems to us particularly meaningful.

> "*Since the nineteenth century the great majority of scholars have held that ancient science and modern science are worlds apart. But if one reads through the texts collected in the Source Book*[38]*, he can not but agree with the editors that it is an error to date the rise of natural science in the seventeenth century and to consider the Greeks «mere speculators». In mathematics, astronomy and mathematical geography, physics, chemistry and chemical technology, geology and meteorology, biology, medicine, physiological psychology, in all these branches of learning the Greeks developed and followed methods that closely approximate, if they do not equal, the standards of modern science. To be sure, the material assembled in these chapters is mostly outdated. What is presented here is not yet modern science. Nevertheless the link between the ancient investigations and those of modern times is obvious"* [39].

End after

> *"That ancient science failed to lead to technological application is another one of those prejudices that die hard. Yet. contrary to the assertions repeated over and over again and made the basis of far-reaching generalizations, like those of Spengler, the Greeks were not hostile to technology, Plato, to be sure, blamed the «corrupters and destroyers of the pure excellence of geometry, which thus turned her back upon the incorporeal things of abstract thought and descended to the things of sense, making use, moreover, of objects which required such mean and manual labor». But Plato is not all of antiquity. Archytas, Eudoxus, Menaechmus constructed instruments and machines. Aristotle admired mechanical toys. Aristoxenus appreciated technical detail. Although Plutarch intimates that Archimedes on account of his «lofty spirit», his « profound soul», that is, on account of his Platonic leanings, did not write on his inventions it still remain true that this «geometrical Briareus", as the Romans called him, did apply his knowledge to practical ends. The list of his inventions is impressive. Geminus, among others, considered mechanics a branch of that part of mathematics which is "concerned with and applied to things perceived by the senses". ...»"*[40].

But Edelstein himself warns:

> *"Modem science and ancient science, then, are not diametrically opposed. I hasten to add, however, that such a claim can be made good only so long as one is willing to do what the editors of the Source Book have done, namely to select as evidence that material «which would generally be regarded today as scientific in method, i.e., based, in principle, either on mathematics or on empirical verification». To put it differently, the impression that ancient science is modern in character is bought at the price of neglecting or omitting all the evidence to the contrary*"[41].

Admitting therefore that the things are exactly as described, even with different tones, two great problems are set.

1. Why, even in presence of enough theoretical premises, a scientific-technological civilization, as that started in that age, would not have been developed? Obviously this question doesn't subsist if we assume with Russo[42] that at least the start of such civilization would take place.
2. Why the initial progress of the Hellenistic civilization in brief time was stopped, declining with great rapidity?

---

[38] The considered article was written by Edelstein about the book of COEN, M. R., 1948;
[39] EDELSTEIN, p. 91.
[40] Ibid., pp. 96-97.
[41] Ibid., p. 93.
[42] RUSSO, 1997. (See also at sect 1. Introduction)

The answers are obviously different. Edelstein write:

> "*The argument most commonly advanced, and advocated also in the now most widely read histories of ancient science, is that in a slave society labor is cheap; technical improvements therefore were unnecessary in antiquity. Such an oversimplification seems no longer justifiable, for as the investigations of the past few decades have shown, ancient economy can hardly be called a pure slave economy. Especially in the arts and crafts free labor continued to hold its place. In addition to slaves, metics and citizens were employed as artisans during the classical age; .... Simple reference to ancient society as a slave economy, then, explains nothing. The exact numerical relation of the various components of the laboring class it is difficult to estimate. What is certain is that the number of slaves in antiquity was much smaller than was thought by historians of the nineteenth century , and this is true above all of the classical and Roman ages. However, even assuming that the percentage of slaves was relatively high, slave labor was neither cheap, nor docile, as is evidenced by slave revolts and strikes.*"[43].

After having criticized the positions founded in exclusive way upon economical evaluations, Edelstein continues, in pressing opposition to Benjamin Farrington[44], of which disapproves the two fundamental theses, that is that the missed scientific development in the antiquity would essentially be owed to two factors: 1) An ideological refusal of the technology; 2) The censorial intervention of the politics. The reasoning of Edelstein is convincing, historically found upon valid data, and it doesn't seem to leave space to meaningful objections. But when he tries then to answer to the same questions, he thinks to found the causes that braked the scientific-technological development in the individual and private character of the search and in the lack of an organized scientific enterprise. To show this he uses data and sources that originate or from the classical period or from the following Greek-Roman one, with exclusion of the Alexandrine period: the only one, in the whole antiquity, in which the scientific search is surely organized and enjoys of conspicuous financings. The Library and the Museum of Alexandria, other smaller institutions elsewhere existing, the realizations of Archimedes, are certainly known and recognized, but they are considered few influential exceptions. It is evident that once more there is a refusal to hypothesize a clean cut and a sudden arrest that could had place in the following period, refusal that is present in the majority of the modern researchers. We think to see on such refusal a positivistic idea of progress, as a continuous and unstoppable *to go forward*. According to this idea, if some scientific and technological progress was started then it can't was stopped without a very big exceptional and traumatic event. It is just in opposition to this, that the book of Russo was inserted here. He sustains in fact that the scientific revolution of which he speaks, would be developed in the brief arc of time characterized by the economic, military, politic and cultural power of Ptolemaic Egypt and of the other Hellenistic Kingdoms, becoming unintelligible after the Roman conquest. In such way most of the possible objections is overcome, because all the contrary results, that we have, concern preceding or following periods, or they are extraneous however to the cultural elaboration that is developed in the Alexandrine area.

4. Interpretation of the science and the σωζειν τα φαινομενα.

The definition of such an ample problem, introduces not only great difficulties for the shortage of original documentary material[45], but it risks to give disputes that are *empty* of their object.

---

[43] Edelstein., pp. 97-98.

[44] See FARRINGTON, 1944.

[45] Almost all the philosophical texts and literary products of the third century B.C. are lost, while the scientific texts that remain, among which also some of the most important ones, are generally without comment and above all they don't declare the philosophical choices set to their base.

This because the different sustained theses are often reported to different conceptions of the science. Already more above we quoted a passage of Edelstein in which this difficulty comes to delineate in enough evident way (see quotation corresponding to note 43) if it is put in relationship with what Edelstein himself says later:

> "*The editors* [of the book. See. note 38], *like many other students of antiquity, seem inclined to classify «theories that are now known to be false or even ridiculous» as «magic, superstition, and religion.» They speak of «'pseudo science,' such as astrology and the like,» that «can be found in the writings of such sober Greek scientists as Aristotle and Ptolemy»; they refer to «the intrusion of the occult» that is noticeable also in modern scientific writings from Kepler to Eddington. But astrology, the theory of humors, Plato's mathematical scale of music are not «intrusions» in ancient science. Theories like these, which do not pass the muster of modem criticism, constitute is in fact the greater part of the preserved material. To the Greeks, they were, just as scientific as those other views which happen to seem acceptable to the modern scientist"* [46].

It is nevertheless very unlikely that this judgment can be adapted to the scientists of the Alexandrine period as Euclid or Archimedes and Apollonius, also because, as we already said (See note 45), in their scientific texts there are not beliefs and visions of the world. The reference to Ptolemy, lived in the II century B.C., doesn't add instead nothing to how much already said. Nevertheless the quoted passage clarifies in effective way the problem to which we want to refer and that often makes not comparable among them the evaluations on the ancient sciences, just for the incommensurability of what is intended by the word "science" [47].

One of the more live debates on the interpretation of the Greek science, in general and not only of that one Alexandrine, concerns the saying σώζειν τὰ φαινόμενα (to save the phenomena), used particularly by Pierre Duhem[48] to reinterpret the most meaningful part of the ancient science in instrumentalist key[49].

We agree with a big part of the criticisms addressed to Duhem when, forcing the sense of the words or distorting a translation, he attributes not demonstrable instrumental intents to the "*most representative"* of the ancient Greek scientists and in particular way to Ptolemy and Proclus. Nevertheless, Geoffrey Lloyd[50] himself, showing many of the historical mistakes of Duhem, warns against easy generalizations. He says in fact in the introduction of his "To save the phenomena":

---

[46] EDELSTEIN, p. 93.

[47] It is hardly the case to remember as a large part of the post-Popperian criticism (among which T. Kuhn, P. Feyerabend, etc…) conducted to more and more vanished and mobile vision of the delimitations between science and '*pseudo science'*. In this perspective, also the conceptions, that are introduced in the history as soaked of elements of metaphysics or also of "*occult*" or "*magic*", could not immediately and uncritically be liquidated as '*superstition*' and '*pseudo science*'.

[48] DUHEM, 1908. See also DUHEM 1956-73.

[49] By the word *instrumentalism* we understand an attitude that sees the scientific theory as an artificially built *tool* to insert the observed data in a coherent system with the purpose to allow predictive deductions on the future phenomena to check experimentally. *Realism* is the opposed attitude to the *instrumentalism*, that is a tendency to consider the scientific theories as real explanations of reality. On this matter, as it is known, the process to Galileo a lot was played, because the Cardinal Bellarmino was well inclined to accept a *purely mathematic* description of the cosmological system in terms of a *heliocentric model*, to the condition that the conviction that "*really* the sun is immovable to the centre and the earth is moved" was repudiated, while Galileo was well firm in his realistic position. Duhem reinterprets the whole history of the physics and above all of the cosmology in terms of *to save the phenomena*, understood in the sense of a substantial and, often, radical instrumentalism, so he reach the conclusion that the positions of Bellarmino were correct and wrong the realistic one of Galileo. Here however we are interested to the fact that the instrumentalist interpretation of Duhem is extended to the greatest part of the scientific elaborations of the Greek antiquity also extended to astronomers as Ptolemy.

[50] See for ex. LLOYD, 1993.

> *"First of all, the pluralism of the ancient science must be remarked again.* […] *the purposes and the assumptions of the ancient scientists in the field of the astronomy, of the acoustics, of the optics, etc…, is different, and not only from discipline to discipline, but also inside every of them. So, inside the same astronomy there are different undertaken types of study or types of composed essays. Nearby to the tradition represented by the construction of mathematical astronomic models by Hipparchus and Ptolemy, there is, on the one hand, a more descriptive work search and, on the other hand, a work of more mathematical character* [...]. *I cited the tightly geometric study of Aristarchus «On the dimensions and the distances of the sun and the moon»: such tradition also includes the «Spherical» of Theodosius, the writing «on the mobile spheres» of Autolycus and the «Phenomena» of Euclid"* [51].

And even if in the conclusion of the same essay, he affirms that:

> "*there where in effect we have some documentations,* […] *they often contradict the interpretative line so emphatically sustained by Duhem and then taken back by others.* […] *In the methodological declarations of Geminus, Theon and Proclus and in the real scientific practice of Ptolemy we find elements in support to the opposite point* [of view]" [52],

he premises nevertheless that:

> *"…for a lot of the most important figures of the history of the Greek astronomy we are not in condition to pronounce ourselves in a definitive way on their conceptions or on the status of the varied hypotheses from them used or on the more general matter of the nature of the astronomy and of his relationship with the physics"* [53].

Now this last observation unfolds all its meaning if we reflect on the circumstance that really for Euclid and for other scientists of the considered period, the lack of knowledge regarding their philosophical beliefs is not owed, as in other cases, to the loss of their works, but to the fact that they wontedly have omitted every pronouncement[54]. We will take back forward this problem that seems to us essential because we don't think that the apparent to be aseptic of the scientific writings of Euclid and other coeval scientists can be due to the chance or to an absence of convictions; neither the fact can be casual that, contrarily to a few centuries from there, all those persons that will attend to mathematics and exact sciences, will remark the exigency to accompany the scientific text with ample comments and justifications of the chosen premises. It is at least the sign of a conceptual change, or, to saying it with Kuhn, of a "Gestalt reorientation" [55] towards the methods and the objects of the scientific '*to know'*.

About the different conceptions of the science, we must say that Russo, in his mentioned book[56], specifies in precise way what he intends with the expression "*exact sciences*", preciously:

1. *"...* [they are] *constituted by theories, as Thermodynamics, the Euclidean Geometry or the Calculus of Probability, with the followings main characteristic points: The scientific affirmations don't concern concrete objects but specific theoretical beings* […].
2. *The theory has a rigorously deductive structure; it is constituted, that is, by few fundamental propositions («axioms», «postulates» or «principles») on his own characteristic beings and from an unitary and universally approved method to deduce a*

[51] Ibid. p. 427.
[52] Ibid. p. 470.
[53] Ibid. p. 470.
[54] Nevertheless, as we will say forward, the "Phenomena" of Euclid are not entirely deprived of indications toward a mathematical model that «*save the phenomena*».
[55] The Gestalt psychology considers ambiguous figures that can be interpreted in different way in different moments. Typical is for ex. the case of geometric figures that appear sometimes concave, sometimes convex, according to the mental disposition of the observer; disposition that can suddenly change without there is apparently a reason. This "*mutability*" of gestalt's orientation is assumed by Kuhn as metaphor of the change, not only of individual level but, for vast cultural areas, of a deeper and general vision of the world. (See for ex. the intervention of Kuhn in "Criticism and growth of the knowledge", edited by Imre Lakatos and Alan Musgrave, Italian translation, Feltrinelli, Milano, 1984).
[56] RUSSO, 1997.

beings and from an unitary and universally approved method to deduce a boundless number of consequences. In other words the theory furnishes general methods to resolve an indefinite number of problems. Such problems, enunciable within the structure of the theory, are in reality «exercises»: problems, that is, on which there is a general accord among the experts on the methods that can be used to solve them and to check the correctness of the solution. The fundamental methods are the proofs and the calculus.

3. *The «truth» of the «scientific» affirmations is therefore in this sense guaranteed. The applications to the real world are founded on the «rules of correspondence» between beings of the theory and «concrete objects»"*[57].

We don't enter into a discussion on those we think to be the limits of a definition of the science which is expressed from Russo[58], also because to the purposes of the present paper, we don't believe that a complete answer is necessary to all the themes that we placed until now, neither we think necessary that a general characterization of the Alexandrine-Ptolemaic science is given. Euclid, in fact, could participate only to the initial phase of the Hellenistic age. Instead it seems to us very important to consider some of the original characters of this period that we can't anymore found in following phases of the Hellenistic age, because the philosophies and the visions of the world that was the base of the scientific progress in this century will be subsequently refused. It is necessary however to don't fall into temptation to look for an unitary and typically Alexandrine vision of the world, rather it is just this lack of unitary visions and values that seems to be the real character of the century and it is just here that we must look for a key of reading of the character that the different sciences begin assuming.

In the complex and variegated differentiation of the points of view that are faced, often in sour polemic among them, it is an obligation to make reference to those are considered the three main current of the philosophical post-Aristotelian thought: the Epicureanism, the Scepticism and the Stoicism even if we have to say that this is only a simplification for convenience. So, for example, we cannot ignore the contemporary existence of reality as the peripatetic school, but it doesn't seem that the immediate followers of Aristotle has product something of meaningful for us. In the same way we have to consider the differentiations, also radicals, that divide philosophers that was too easily confused by the use of a same label as Sceptics[59], Stoic, etc….

Having therefore to make reference to the philosophical currents of the third century B.C., and specially to the relationships that can subsist with the mathematics and the exact sciences, we want here to contract the discourse to the Stoic school founded from Zeno of Citium and to the Sceptics of the Academy beginning from Arcesilaus of Pitane. This choice is owed to the fact that the contrast of ideas between these two schools seems to us of particular interest; contrast

---

[57] Ibid, p. 34.

[58] This is not the place to express our position on the epistemological criticism that in the XX century followed the logical positivism. We believe to be useful only to develop some particular considerations about this specific case. It seem to us that putting a too sever limitations to what is legitimated to call science (as Russo do), while on the one hand can be useful, helping to separate confused things among them, on the other hand risks to artificially build a cut, so rigid to lose sight of meaningful parts of the complex historical circumstances. A sure merit of the book of Russo with his definition of exact sciences, is to delimit in a precise way the problem of the validation of the historical sources. So, putting also a precise space-temporal delimitation, he starts using improper generalizations to which we have still now more times mentioned (see in particular the end of sect. 3). Oppositely a rigid delimitation as we have seen, of what is right science, would too drastically exclude visions and systems as the Aristotelian one that, if on the one hand, with his pretensions of metaphysical absoluteness, constituted an obstacle to the empirical investigation, on the other hand has been and it is a base that keeps permeating of itself also the more positivistic regions of the modern and contemporary science. (See note 27).

[59] We remember as Arcesilaus held a lot to specify a connection of descent with Plato, to distinguish in clear way from Pyrrhon and from the Phirrhonian scepticism.

that during the dialectical opposition, first between Arcesilaus and Zeno, more forward between Carneades and Crysippus, produced certainly for both them a growth of depth of elaboration[60].
We don't think, as already said, that direct relationships of adhesion can be found, and this for a multiplicity of reasons both of chronological character, and for the already many times mentioned aseptic character of the scientific texts. It appears well funded, instead, the problem to verify, through the comparison with the different expressions of the coeval thought, what hypothesis on the Greek science can be considered somehow compatible with other aspects of the culture of the same age. It is enough for us, to point out as the foundation of the Hellenistic Kingdoms constitutes an element of strong novelty because:

1. The cultural production loses its purely individual volunteer and private character, to become financed activity, organized and integrated in the government structure.
2. Between science and technique there are interrelations that are not casual but surely required by the same government entity (the king) that finances the production and diffusion of the culture. It doesn't care, for the time being, how much wide was their economic importance and if the prevailing aim was of military type.
3. The Hellenistic world is from now so complex and variegated that, in it, an unified vision of the world cannot be anymore thought. The science, intended as complex of particular sciences that can have some relationship with the τέχνη, can subsist then only under the condition to don't express any opinion on the themes of the *being as being* and to leave the matter of the absolute and definitive truth of the scientific premises undecided[61].

If the first two points seem already enough clear and hardly disprovable, some precise statements can be useful on the point 3.
If, for one hand, it is difficult to maintain a precise interest towards the particular sciences by the greatest representatives of the philosophical Hellenistic schools[62], we are not able nevertheless to deny that the new terms in which the philosophical debate is set, putting in discussion the same meaning of knowledge, can't not have more direct consequences on the methodologies and on the bases of the scientific search. The simple distinction, for instance, within the Stoic school, between reality and meaning of the discourse (λεχτόν), sets in absolutely new general terms the problem of the knowledge and particularly to the scientific knowledge. Scepticism and Epicureanism, agree then in to assign to the phenomenical experience the only source of knowledge that we can have, even if only the first one between such two schools assigns a value to the deductive method. The most serious problem, however, from the point of view of the value that one wants to assign to the scientific knowledge, it is in the effective stability and objectivity of the possible knowledge. It is this one the point of real and apparently irremediable contrast among Stoics and Academicians. For the Stoic ones the science (ἐπιστήμη), differs from the opinion (δόξα) because the first one have reached such a stability that cannot be upside-down by reasonings.
We can wonder "why they specify «by reasonings»?". It is possible to think that they considered the possibility to change some previously admitted knowledge by a different way? For example if a new comprehensive representation (χαταληπτιχὴ φαντασία) arrives as consequence to some new experiences? All the interpretations given by the ancient sources and by the modern exegetic researches, lead to a negative answer. But we have to consider also that the Stoic phi-

[60] See in particular IOPPOLO, 1986.
[61] On the Hellenistic history and society, as well as on the organization of the scientific enterprise in Ptolemaic Egypt, see for ex. BERGTSON, 1989; BIANCHI BANDINELLI, 1977; FRASER, 1972; GULLINI, 1998, FLOWER. For the philosophical thought in the same period, and particularly for the dispute among Stoics and Academicians, see for ex.. GEIMONAT, 1973; IOPPOLO, 1996; ISNARDI PARENTE, 1994, 1999; LÉVI, 2002; CANFORA, 1995.
[62] See for ex. the introductive essay in ISNARDI PARENTE, 1999).

losophers asserted, of course, the possibility to become a wise man, but nobody says to be a wise man, so such possibility has to be considered as an ideal target.
There is however a point that perhaps could produce a doubt, and however can show as the ancient Stoic gnoseology was more complex that it seems. We find it among a fragment of Galen, the author that, just for the passionate vehemence by which he expresses his aversion for Crysippus, refers, to testify his objectivity, some integral passages of the Stoic philosopher. We see in fact that according to Galen[63]:

> "[Crysippus] *undertakes to show what is correct to believe on the ground of the opinion* [δοχα] *of any testimonies of the common people and not according to the nature of the things"* [64].

Obviously this is by itself few believable because in clear contradiction with the central nucleus of the Stoic doctrine and particularly that one of Crysippus. But Galen continues:

> *"I transcribe here his same expressions, that are about these: «about such things we will likewise make search, starting from the common opinion and from the discourses that according to this taken place» and with common opinion Cryisippus wants point out what commonly appears to all people; then continuing he says: «by all of this, since the beginning of preference, they seem to be conducted to affirm that our directive part is in the heart». Still treating more of this, he textually writes: «it seems to me that most men are generally inclined to affirm this, because in certain way they realized that, in concomitance with their psychic motions, it is verified something in their breast, and especially in the place where the heart is set…»"*[65]

And so Galen polemically describes the defence that Crysippus is building to his thesis; but more than the matter on which they contend we are interested to the way by which the Stoic philosopher attributes cognitive validity to the common opinions, to the tradition, to the myth, to the allegory (Galen say sarcastically :"…*After to have filled the whole book with verses of Homerus, Hesiodus, Stesicorus, Empedocles, Orpheus*,…."[66]). Still in the testimony of Galen, we find however in another point (on the matter if the heart was the source of the nerves) the admission by Crysippus …*of do not know really the matter because inexperienced of dissection*. Precisely Galen says:

> "[…] *Nevertheless he makes tolerable his opinion by saying modestly that he doesn't suppose to say that the heart is the source of the nerves or that he knows really what is concerned to this matter, since it is pronounced inexperienced of the art of dissection."* [67]

These last affirmations, that Galen interprets as proof of a confessed incapability and incompetence of Crysippus, can be interpreted instead as awareness that the cultural stratifications bequeathed by the language, the tradition and the myth, contain in itself an image of the world that possesses its own validity, surmountable[68], as it seems, only by a reorganization of the knowledge founded on specific and organized experiences (in this case the dissection). And in fact, if Crysippus declares to don't know the reality on this matter, to what would be directed the long discussion on which Galen referred, if not to the interpretation of the symbolic meanings of the myth and the semantic stratifications of the language, as another passage of Crysippus, referred

---

[63] The theme of the dispute concerns the location of the *directive part of the soul* that is, in the ancient language, the location of the our rational functions. The fact that Crysippus defends the most retrograde thesis (that it is in the heart) has here a little importance: we are interesting only to some reasoning's passages.
[64] GALENO, *De Hippocr. et Ptat. plac.,* III, 1, p. 254 Müller = SVF II, 886 . Cit in ISNARDI PARENTE 1999, p. 400.
[65] Ibid.
[66] Ibid., III, 4, p. 281, Müller = SVF II, 907 . In ISNARDI PARENTE, p. 413.
[67] Ibid., I, 5-10, p. 138-145, Müller = SVF II, 897 . In ISNARDI PARENTE, p. 407.
[68] That the tradition, as the language and the myth can't be, for Crysippus, the final criterion of the Truth, is clear for many others known fragments. In particular we remember that he wrote against the common opinions.

by Galen, confirms?[69] In particular way just the sentence of Crysippus "*besides we give birth in us to the products of the sciences*" seems to oppose in radical way to the tightly realistic interpretations given by Galen[70].

Obviously this doesn't oppose with the fact that the Stoic philosophers aimed to reach a definitive and stable knowledge. Here are the fundamental terms of the opposition between Dogmatists (Stoics) and Academicians (Sceptics), controversy whose fundamental nucleus concerned the possibility to reach a stable and objective knowledge. Such possibility, as we said, was admitted by the Stoic philosophers, at least as principle, but as result of a long process of intellectual elaboration beginning from the data of the sensitive experience. It is here however essential the fact that the sensitive datum is not the knowledge by itself, because this passes through the creation of conceptual objects or categories of thought (προλήψεις ο κοιναί ἔννοιαι) that are expressed and perhaps partly identified with the ways of the language. So the *logos* becomes essential part of the rational knowledge which begin seems to be set about the seven year-old age[71].

The Sceptical Academic school, instead, that began when Arcesilaus becomes chief of the Academy, contested not so much the most complex procedures by which the knowledge should be acquired, but the idea in itself that a certainty on something could be however reached. They advocated therefore the necessity of the suspension of the judgment, but not for this reason they refused the *practical* forms of knowledge, intended as provisional acquisitions, always revisable and revolts to some finalities (τέχνη). It is prevailing, but not unanimous[72], opinion that both the schools didn't look with interest at the particular sciences, but only to the ethical matters. What, however, is interesting for us and appears as a fact, is that the general frame was surely not incompatible with a more and more autonomous, and somehow *not realistic*, development of the particular sciences. Rather we could say that the ideal frame for the development of scientific ideas was that one furnished by the philosophical disputes on the value of the knowledge, as those that took place especially between Stoic and Academic philosophers. A frame in which the science could earn authoritativeness assuming new logical dimensions and categories of thought, at least similar to those ones developed from the Stoic school[73], but also suspending the judgment, as the Academicians wanted, on the reality of the things and on the truth of the principles.

---

[69] "*After this, Crysippus* [...] *says: «Such they are the things that are said of Athena* [the myth that wants Athena having birth by the head of Zeus after that this swallowed Metis]*, and* ***the allegory that results from them*** *is another. For first thing Metis is compared to the mind and the art of to live; for artwork of this we have to send down and to swallow the arts equally that we say to send down discourses of other peoples: it is consequently as to say that we almost have to gulp down them and to send down them in the abdomen. After this, it is reasonable that we give birth to this art that we swallowed, becoming with this similar to a mother that produces; besides we give birth in us to the products of the sciences...»*" (Ibid., III, 8, p. 321, Müller = SVF II, 909 . In ISNARDI PARENTE, p. 415).

[70] However all the testimonies on Crysippus, to a careful reading, bring to the representation of an extremely deep thought, beyond, for the most part, than the same witness not succeeded in intending.

[71] "*Just that ability to reason in virtue of which we are said reasonable beings, they says that it is formed in us in base to the anticipations (προλήψεις) and it reaches perfection around the seven year-old age"*. (Aetius, Plac., II, 1-4, Dox., Gr., pp. 400-401 =SVF II, 83, in ISNARDI PARENTE, p. 700). We notice transiently only the almost coincidence of age, in the theory of Jean Piagét, with the passage from the pre-operating intelligence to that of the material operations. See for ex. FLAVELL, J. H., *The developmental psychology of Jean Piaget*, Princeton, N. J. , **1963**)

[72] See for ex. LÉVY.

[73] As it regards the revaluation of the stoic logic, not all arrange on the value that it could have in relationship to the sciences. Isnardi Parenti thinks, for instance, that the logic of Crysippus constitutes, in comparison to the Aristotelian one, a return to the tradition of the fifth century. This particularly because the Crysippus's syllogism, unlike the Aristotelian one, have as object individual and not universal things. We don't agree on this, and observe that in the geometric proofs the form of the Aristotelian syllogism was never used. The reasoning in fact is always referred to singular objects, even they are supposed selected in a casual way. We say for instance: the point A.…, the point B, the straight line AB…., the angle ABC…, etc…; never "all the points are…" or "all the straight lines are….", etc…The generalization is obtained to a next action of thought in virtue of the arbitrariness with which the choices are intended effectued. Syllogisms of the type "*All the triangles are polygons, all the polygons are surfaces, therefore all the triangles are surfaces*", can help us to illustrate with examples the Aristotelian syllogistic forms, but

5. Common notions and postulates.

A datum already meaningful for itself, is the fact that the authenticity of the *common notions* has been debated, in everything or partially. Putting aside the recognized unreliability of some of them, we think interesting the debate to which an article of Tannery given place, by which he sustained a not authenticity of all the common notions, that would been subsequently interpolated, and he formulated the hypothesis that Apollonius would have add them. These conclusions have been however rejected by Heath and by other authors. Even if there are not certain documentary proofs, we are inclinable to accept the prevailing thesis, according to which at least the first three common notions (but probably more) would be authentic of Euclid. Any the factual truth would be, nevertheless, we consider very interesting the reasoning produced on both sides[74]. To such purpose we entirely report here what Heath said on the subject:

> *"The following are his main arguments. (1) If Euclid had set about distinguishing between indemonstrable principles (a) common to all demonstrative sciences and (b) peculiar to geometry, he would, says Tannery, certainly not have placed the common principles second and the special principles (the Postulates) first. (2) If the Common Notions are Euclid's, this designation of them must be his too; for he must have used some name to distinguish them from the Postulates and, if he had used another name, such as Axioms, it is impossible to imagine why that name was changed afterwards for a less suitable one. The word ἔννοια (notion), says Tannery, never signified a notion in the sense of a proposition, but a notion of some object, nor is it found in any technical sense in Plato and Aristotle. (3) Tannery's own view was that the formulation of the Common Notions dates from the time of Apollonius, and that it was inspired by his work relating to the Elements (we know from Proclus that Apollonius tried to prove the Common Notions). This idea, Tannery thought, was confirmed by a «fortunate coincidence furnished by the occurrence of the word ἔννοια (notion) in a quotation by Proclus (p. 100, 6)*: *"we shall agree with Apollonius when he says that we have a notion (ἔννοια) of a line when we order the lengths, only, of roads or walls to be measured.»*
>
> *In reply to argument (1) that it is an unnatural order to place the purely geometrical Postulates first, and the Common Notions, which are not peculiar to geometry, last, it may be pointed out that it would surely have been a still more awkward arrangement to give the Definitions first and then to separate from them, by the interposition of the Common Notions, the Postulates, which are so closely connected with the Definitions in that they proceed to postulate the existence of certain of the things defined, namely straight lines and circles. (2) Though it is true that ἔννοια in Plato and Aristotle is generally a notion of an object, not of a fact or proposition, there are instances in Aristotle where it does mean a notion of a fact: thus in the Eth. Nic. IX. 1171a32 he speaks of "the notion (or consciousness) that friends sympathise «(ἡ ἔννοια τοῦ συναλγεῖν τοὺς φίλους) and again, b 14, of "*

---

they are too much trivial to be able to do indeed geometric proofs. If instead one had wanted to use however the universal quantification (as Aristotle would like) in the real mathematical proofs, he would necessarily have had to bring the actual infinite to the account. This is so the problem that, with the development of the modern infinitesimal analysis, has not been more possible to elude and that has brought to the set theory. In terms of modern formal logic, this is gotten adding to the Modus Ponens the Generalization rule, that affixing to a variable the universal quantificator (for all), generalize singular propositions to an actual infinity, whereas supposing only the arbitrariness of the choice (as in the case of Euclid, but of Cauchy also), the extension becomes only potential. In conclusion we think important, in the evolution of the mathematical thought, the Aristotelian conception of the syllogistic reasoning as formal tool separated by the questions of truth and reality. If the real form of the syllogism is considered instead in the mathematical reasonings, then that of Crysippus appears to be more qualified than the Aristotelian one.

[74] TANNERY, 1884.

the notion (or consciousness) that they are placed, at his good fortune.» It is true that Plato and Aristotle do not use the word in a technical sense; but neither was there apparently in Aristotle's time any fixed technical term for what we call «axioms», since he speaks of them variously as «the so-called axioms in mathematics,» «the so-called common axioms,» «the common (things)» (τὰ κοινά), and even «the common opinions» (κοιναὶ δόξαι). I see therefore no reason why Euclid should not himself have given a technical sense to «Common Notions,» which is at least a distinct improvement upon «common opinions.» (3) The use of ἔννοια in Proclus' quotation from Apollonius seems to me to be an unfortunate, rather than a fortunate, coincidence from Tannery's point of view, for it is there used precisely in the old sense of the notion of an object (in that case a line)"[75].

We reported this long passage because some meaningful elements emerge from it for our analysis. But we want first to recall on this subject the judgment of Enriques that affirms:

"*The distinction between the postulates and those that forward are designated as «common notions» (the «axioms» of the Pythagoreans) is illustrated by Aristotle and by the comment of Proclus according to different points of view* […]"[76]

and after:

"*But it is remarkable that Aristotle never speaks of common notions, using the Pythagorean term of axioms* (ἀξιώματα = *dignity*); *rather the word* ἔννοια ***doesn't seem to be in technical meaning in Plato or Aristotle****, but only later among the Stoics. However, the deductions that someone (Tannery) wanted to draw from this circumstance, doubting the authenticity of the Euclidean notions, fall in front of the observation that the word* ἔννοια *is found in a fragment of Democritus. And since among his lost works there is an essay of geometry that, for the disposition, is like to the Euclid's Elements, it is permissible to deduce that the text of Democritus could have this denomination of the axioms and that from it Euclid has taken back"*[77].

The reasoning of Heath can indeed frustrate those of Tannery, and we agree with him when he says "*I see therefore no reason why Euclid should himself have given a technical sense to* «*Common Notions*» ". Of course! *Why should not he have does it*? And we tell it beyond the use that Aristotle does in different contexts of expressions as τά κοινα and κοιναὶ δόξαι, without to notice the exigency to fix a technical term. What instead seems to us to unite the three authors is the fact that all are moved on an implicit common assumption that is present in almost all the critical literature on the *Elements* and that can be resumed in this way: "*The Euclidean geometry is built within the Platonist-Aristotelian philosophy; therefore if something in it is not consistent with Plato or Aristotle, and only in this case, or it is a defect or it is a posterior addition*". This assumption his not proposable, we think, because it tends to prove the affiliation of Euclid to a philosophic area by presuming such affiliation as premise. To great reason this holds

---

[75] EATH, 1956, p. 221.

[76] ENRIQUES, F., 1912, p. 42. We remember that really Aristotle distinguishes among the fundamental principles (not demonstrable), those that are to the base of single particular sciences from those that have instead a general character. He however doesn't seem to use for this a technical terminology, as it is noted above by Heath. Terminological distinction was instead done by Aristotle not among the indemonstrable premises but among those that are assumed without proof even if they are demonstrable. Such propositions are said by Aristotle hypotheses if they are believed by the pupils, postulates in the other cases (See Post. Anal. I, 10). What seems to us important is that both of such two order of distinctions have a character instrumental and methodological but not gnoseological. The first distinction (between fundamental or particular principles), in fact, is referred to the generality degree, the second distinction (hypotheses or postulates) concern a precise rapport with well determined pupils (as it is well marked by Aristotle) that can or not believe some premise. As we can see, all this do not regard any problem of being and of truth. To the first principles (indemonstrable) one riches, for Aristotle, following an inductive process and a not better clarified intuition; therefore, to be such, they have to be as much clear, as evident and true.

[77] Ibid. pp. 47-48.

if we consider the complexity of the cultural, scientific and philosophical context of the third century B.C., during which, on the other hand, we don't known what of the Aristotelian works were known and circulating[78].

Going back to the quoted passage of Enriques, we note as, despite he had observed as the word *ἔννοια* was used by the Stoics (and Zeno of Citium is contemporary of Euclid), he preferred to resort to vague and captious hypotheses, also for the temporal distance that separates Euclid from Democritus. In the Stoic school, not only the word *ἔννοια* is used, but we finds agai the whole expression κοιναί ἔννοια, that for example Isnardi Parente thinks to be able to identify (at least from some citations of Arrianus and Plutarchus) with προλήψεις (translatable as "*idea*" but also "*scheme*" or "*mental image*"; today we would perhaps speak of *cognitive structure*. See also section 4, note 71). We want immediately to observe as the Stoic concept expressed by κοιναί ἔννοιαι, from the extensional point of view could partially correspond to what Aristotle sometimes designates with αχιοματα, but on the conceptual ground differs deeply from it. From here, however, we don't think that we can draw hurried conclusions because, as we already told on introduction, we reject rigid relations of interdependence between Euclidian geometry and Stoic thought[79]. We think, however, that we will must consider with a lot of attention the hypothesis of repositioning the work of Euclid from the narrow circle of the Platonic-Aristotelian philosophy to a more mature phase of evolution of the scientific and philosophical thought. If the knowledge's procedure, in fact, was already theoretically delineated by Aristotle as result of an inductive-deductive process, now it's enriching itself on the one hand through the organized work of the scientists, on the other hand through a philosophical dispute that put in discussion the means of the acquisition and the same value of the knowledge.

Different elements seem to us to point out a coherence with our hypotheses. Meanwhile we must observe that *Postulates* and *Common Notions*, in the *Elements* of Euclid, differ not only for to be separately gathered, but also for the different way of enunciation. Only for the *postulates* (αιτήματα), in fact, the word ἠιτήσθω[80] is initially used; that word is translated by Heath as "*Let the following be postulated"* and by Enriques more literally as "*domanda*" = "*ask*" (as command) while in other cases it is ignored[81]. If we admit then that (at least some of) the Common Notions are authentic, in the sense that they have already been introduced in the original text, then the presence of the incipit ἠιτήσθω in the *Postulates* but not in the *Common Notions*, has of course to constitute a difference between these two classes of first principles. And this difference cannot be confined to the Aristotelian distinction between general principles and principles of the specific sciences, but it has to refer rather to something that involves the state of truth

---

[78] V nota **28**.

[79] The direct approach of Euclid to some aspects of the stoic gnoseology could constitute a very strong and fascinating suggestion. Beyond however of the difficulties that could emerge from a more careful analysis, there is a historical and chronological not eludible datum. In fact the most interesting aspects of the stoic thought in relationship to the theory of the knowledge seem to be those were delineating with Crysippus, not in a casual way, but in the course and under the push of the controversy with the Academy (See. JOPPOLO). Now, because Crysppus borns in 281, it is not chronologically possible that he had some influences on the our Alexandrine scientist, active certainly before he was born. We cannot be exclude instead that the new forms of thought that were affirming in the science could influence the philosophical debate: it would seem rather few believable to us that this had not happened.

[80] Ἠιτέσθω: Imperative form of the verb αἰτέω (to ask), therefore literally «*ask*» as command. In the dialectical tradition of the Greek philosophy the questions were put to the interlocutor before beginning a deductive reasoning and they were substantially a request of assent on some propositions, that after approved became the premises of a reasoning. The same Aristotle sometimes uses the term *asks* with the meaning of premise of a syllogism. It is clear from this that the intent is to constitute a syllogistic premise, but it doesn't automatically follows from this, that such premise has a hypothetical value in the sense of suspension of the judgment on the truth. Just for what the comparison with the common notions, can assume meaning in this sense.

[81] See for ex. TRUDEAU.

or any other criterion of acceptability. Now, because the dialectical question is just a request of assent, the different way of setting the two classes of first principles can be formulated as follows: the assent of the reader was trivially admitted for the *Common Notions*, while for the *Postulates* a convention, an agreement, or a hypothetical acceptance was required[82].

We want besides to do a last remark on this argument: as we said (see note 76), the term *postulate* had for Aristotle a very different meaning, at least if we refer ourselves to the *Posterior Analytics*. In particular it is impossible to assimilate the Euclidean distinction between *postulates* and *common notions* with the Aristotelian distinction between *postulates* and *hypotheses* because: a) the Aristotelian division is (as he explicitly remarks) relative to particular pupils and so it is possible only on the personal rapport of teaching and not in a write work. b) Some of the Euclidean common notions, is used by Aristotle as example of *indemonstrable general premise* that he distinguishs from both of *hypotheses* and *postulates*.

But beyond this indication, there are deeper motives to think that Euclid, on the matters of truth and existence, had a different position than Aristotle. We will consider in the next section the most meaningful topic under this aspect, that is that of the parallel straight lines. Now we want instead to remember that Euclid, besides the *Elements*, also wrote other works and, among these ones, the *Optics* in particular must be considered, just because they reflect the same hypothetical-deductive structure of the *Elements*. Here also the enunciation of the fundamental principles (Postulates) begins with a hypothetical form, although in this case the used term is different[83]. But what is here more meaningful is in the content itself of the postulates, in particular for the postulate 1, 3 and 4 that are formulated as follows:

1. *Let us to put therefore that the radiuses[84] that depart from the eye are straight lines that have some distance among them.*

……..

3. *and that the things to which radiuses come are visible.*
4. *and that the things to which radiuses don't come are not visible.*

It is clear that the criterion of the truth and the self-evidence prescribed by Aristotle to accept the prime principles, is here totally inapplicable, and if one wanted to adopt it, than it would bring only to the refusal of the theory; and really this historically happened[85]. In fact, also without giving a realistic interpretation to the radiuses (see note 84) and considering only them as geometric entity, it would be difficult to find a justification to the choice of a discreet model (the *radiuses* have distance among them) considering only the axioms in itself, neither the reasons because the

---

[82] Obviously all this would fall if one accepts, in accord with Tannery, the non authenticity of **all** the common notions, and in this case a possible stoic origin of their denomination could be reconsidered. But the idea of Tannery that the interpolation of common notions can be attributed to Apollonius, beyond the reasons produced by Heath, appears to us not proposable because, the custom of the comments and of the interpolations in other people's works seems begun only later, when the original scientific production become less meaningful and, above all, the scientist coincides often with a philosopher that tray to brought again the science inside some philosophical conception. Securely this condition is realized starting from Posidonius and Geminus in the first century B.C.

[83] ὑποχεί́σθω = *Let us to put.*

[84] In the translation, we used here the term *radius* and not *ray*, because such last one presuppose a realistic interpretation, that is *rays of light*; on the contrary we think, according to Incardona, that it has to be interpreted as geometric straight lines conducted from the eye.

[85] Historically it was refused (for ex. by Ptolemy and by Leonardo of Vinci) by virtue of a realistic interpretation of the "*rays* (or radiuses*?) that go out of the eye*" (See for ex. INCARDONA). Such interpretation kept on influencing the modern criticism also, despite the purely mathematical character of the Euclidean text was clear. We read for example in BOYER (that in turn mention COHEN and DRABKIN): "*The optics of Euclid is notable for the exposure of an «emissive» theory of the vision according to which the eye utters rays that cross the space*…."(BOYER, 1976, p. 121). End immediately later: "*Note that the mathematical concepts of the perspective (otherwise from the physical description) are the same any it is the adopted theory*" (Ibid). Nevertheless, considering that the Euclidean work is only geometric, he don't wonder from where we had to deduce a thought that the author doesn't express. But sometimes the strength of the tradition appears stronger than the observed evidences.

third and the fourth postulates would be included appear understandable if we don't analyse the consequences of those in conjunction with the first one. The things change instead if we assume as criterion of validation not the truth of the principles but the correspondence between the consequences of the postulates and the observable phenomena. In fact, this appears of an extraordinary clarity when one reads the theorems in whose proofs the aforesaid postulates are used. And in fact we find them in the proofs of theorem 2 (Nearer objects are seen with greater resolution than more distant objects) and of theorem 3 (For every object there is a distance over which it is not seen anymore). As we can see, here not only the truth and the self-evidence fails as criterion for the acceptance of the postulates, but the same criterion of validation is moved from the premise to its consequence.

Neither the work *Phenomena* must be neglected to this regard; in fact, even having a discursive style, it displays a notable interest by the point of view that we are examining, just because, as we already said, they are founded upon an initial hypothesis of mathematization that among all the hypothetical formulations of Euclid, is the only one to have been expressly justified in a way that clearly reveals the intent to *save the phenomena*.

6. The question of the parallel straight lines.

On the matter of the fifth postulate we quote by Heath:

> "*We know from Aristotle that up to his time the theory of parallels had not been put on a scientific basis (Anal. Prior. II, 16, 65a 4): there was apparently some* ***petitio principii*** *lurking in it. It seems therefore clear that Euclid was the first to apply the bold remedy of laying down the indispensable principle of the theory in the form of an indemonstrable Postulate*"[86].

We will see the reason why, to our opinion, the theory of the parallel straight lines could not find a coherent organization inside the Aristotelian conception of the indemonstrable principles, neither, all the more reason, in a Platonist philosophy. We will see besides as from here can come down a reasonable key of reading for the two millennia long story of criticism to the fifth postulate. Now however we have to analyze what was said on the subject by Imre Toth[87] in the volume that we have already mentioned in introduction. To the beginning of his work Toth mentions the same passage of Aristotle that few above is cited by Heath, and he says:

> "*The interpretation of a known passage of the first Analytic, II 16, 65a 4-7, reach to the result that this prime principle can be only a proposition that is equivalent or also identical to the famous axiom of the parallel straight lines*"[88].

This passage (of Toth), requires itself some explanation. The quoted passage of Aristotle of which Toth doesn't give a translation, neither precise which is the interpretation to which alludes, is in fact the following:

> _ *ὅπερ ποιοῦσιν οἱ τὰς παραλλήλος οἰόμενοι γράφειν· λανθάνουσι γὰρ ἀυτοὶ ἐαυτὺς τοιαῦτα λαμβάνοντες ἃ οὐχ οἷόν τε ἀποδεῖξαι μὴ οὐσῶν τῶν παραλλήλον.*

There are different discordant translations of it. We quote first the Italian translation of G. Colli that is the following:

---

[86] HEATH, 1931, p. 358.

[87] See. note 9.

[88] TOTH, 2000, p.69. The first principle (αρχἦ), on which Toth spicks and as himself specifies in the same page, is a proposition from which the so-called theorem of the parallel straight lines derives, that is the twenty-ninth proposition of the Elements.

"*Ed è proprio questo l'errore commesso da coloro che ritengono di tracciare delle rette parallele: essi infatti non si accorgono di assumere delle premesse tali, da non poter essere dimostrate, a meno che le rette non si presuppongano come parallele*" [89].

It is a redundant translation that tries to already furnish an interpretation of the text, but of which the meaning is not well understood in relationship to what we know on the problem of the parallels. We have to premise at this point the context in which the passage of Aristotle is found: the philosopher had just illustrated the insubstantiality of the circular reasonings through which some people pretend to prove A through B when B has been in turn proved by C and this by A; the quoted passage is used as an exemplification of these false reasonings. It seems clear, at this point, that the expression "to draw the parallel lines" is referred to some geometric construction of the parallel line drawn by a point to a given straight line; this presuppose some proposition that assures that two straight lines, as are built, are indeed parallel. Toth assumes, as it seems, that the proposition to which Aristotle alludes was the theorem (or equivalent) that Euclid sets as twenty-ninth proposition, the first one in the *Elements* that is proved using the fifth postulate. If so, then indeed a first principle would need equivalent to the fifth postulate to prove such proposition. But what sense does have, in such case, to say that it is necessary to suppose already that the drawn straight lines are parallel? We know that to prove the parallelism between two straight lines drawn in such a way that corresponding angles are equal, needn't the twenty-ninth proposition, but only the twenty-eighth one, which demonstration doesn't require the fifth postulate. A simple construction of the parallel line to the straight line r for a point P is got building first by P the perpendicular t to r (Elem. prop. 12) and then still by P the perpendicular s to t (Elem. prop. 11). By proposition 28 we have finally that r and s are parallel. The twenty-ninth proposition, instead, reversing the proposition 28, allows to prove the oneness (and not the existence) of the parallel line. In this perspective, the Aristotle's passage, in the version of Colli above quoted, doesn't seem to assume a reasonable sense. A rigorously literal translation of the Aristotelian passage could be:

"*This is what those persons do that believe to trace (or to draw) parallel straight lines: in fact, without realizing it, they assume what is not possible to prove if the parallel straight lines don't exist.*"

In this version, that is comforted by other known translations[90], the passage assumes a more coherent meaning; in fact the proposition 28, that allows us to recognize as parallel the built straight line, would not be demonstrable in a geometry in which *the parallel straight lines don't exist*. It is this in fact the case that corresponds to the *hypothesis of the obtuse angle* of Gerolamo Saccheri, in which these three facts, that are interdependent among them, contemporarily hold:

a. The straight line is not infinitely extensible.
b. The sum of the inside angles of a triangle is greater than two right angles.
c. Parallel straight lines don't exist.

What Aristotle seems therefore to affirm, according to what appears the most reasonable interpretation, is that to assure the validity of the construction of parallel straight lines (that is to

---

[89] *"And it is just this one the error committed by those that think to draw parallel straight lines: they in fact do not realize to assume some premises, that can't prove unless the straight lines are not presupposed to be parallel"*. Aristotele, *Organon*, edited by Giorgio Colli, Adelphi, Milano, 2003, p. 250.

[90] For ex. the Italian translation by Marcello Zanatta: "*Cosa che compiono quelli che credono di disegnare le parallele: costoro infatti non si avvedono di assumere cose tali che non è possibile se non esistono le parallele"*[90]. (Aristotele, Organon, a cura di Marcello Zanatta, vol. I, UTET, Torino, 1996, p. 397) or the English one of A. J. Jekinson: "*This is what those persons do who suppose that they are constructing parallel straight lines: they don't see that they are assuming facts which it is impossible to demonstrate unless the parallel straight lines exist*" (Aristotle: *Prior Analytics*, electronic edition of MIT, Massachusetts, on http://*classics.mit.edu/Aristotle/prior.html*).

prove that two straight lines anyhow built, are parallel) it is necessary to suppose the existence of parallel straight lines or some equivalent proposition[91], so if somebody thinks to prove the existence of the parallelism by building parallel straight lines, then he encounters the contradiction considered by Aristotle[92]. From what we said it follows that the required proposition doesn't to be equivalent to the fifth postulate, rather, this cannot be that (or equivalent to that) wanted by Aristotle: in fact the fifth postulate is trivially valid in the case in which the parallel straight lines don't exist. To prove the existence of the parallel straight lines, instead, it is enough a proposition that allows to deny what Gerolamo Saccheri sets as hypothesis of the obtuse angle. In the *Elements* of Euclid such proposition can be identified with the third postulate (unlimited prolongability of the straight line). All of this, to our opinion, doesn't invalidate the general discourse of Toth, but it sets a demand of a more precise distinction between the two fundamental problems regarding the parallel straight lines: the problem of the existence, guaranteed when we declare the unlimited prolongability of the straight lines, and the problem of the singleness, that requires a proposition that must be equal or equivalent to the fifth postulate. This appears clearer, also from Toth, in the introductory essay to the work of Saccheri, written with Elisabetta Cattanei[93]. Here she says in fact:

> "*The hypothesis by which the sum of the angles of a given triangle is greater of two right angles is not by itself incoherent, and it possesses rather remarkable implications of geometric type, however it is incompatible with a fundamental property of the straight line, that is presupposed both by the hypothesis of the right angle, and by the hypothesis of the acute angle: the property of the straight line to be an open in both directions and endless line*" [94].

This is a distinction that, from what can be understood by the quoted passage of Aristotle, didn't have to be still well clear before Euclid, just because a *first principle* had not been found by which it is possible to derive the properties of the parallelism and that appeared, at the same time, simplest and self-evident according to the Aristotelian demand.

To confirm our interpretation of the Aristotelian passage, there are all the other passages of Aristotle that are mentioned by Toth himself. In them in fact, more times and in varied contexts, not Euclidean hypothesis is considered. Every time however such hypothesis doesn't contemplate a geometry of hyperbolic type (sum of angles smaller then two right ones) but, on the opposite, a geometry of elliptic type (sum of angles greater then two right ones) in which cannot exist parallel straight lines. This results is well clear for instance in a passage of the Prior Analytics[95], in which, Aristotle speaks on the contradictions that can subsist after removing a false hypothesis

---

[91] Notoriously in the so-called elliptic geometry, what for ex. is that one on a spherical surface, however two distinct straight lines are given, they always have a common point. If we put ourselves on such geometry, it is possible to draw two lines by the same construction generally used to build parallel straight lines; that is we are always able from a point *P* to conduct the perpendicular *s* to the straight line *r*, and then still from *P* the perpendicular *t* to *s*. In this case however the straight lines *r* and *t* are not parallel because they have a point in common. It is correct therefore to affirm that who believes with the aforesaid construction "*to have built (traced, drawn) the parallel straight lines*" owes first to prove that the geometry is not elliptic, that is that in it the parallel straight lines exist.

[92] Otherwise we don't see where is the circularity in the affirmation that to prove the parallelism between to straight lines, the existence of parallel must be before proved.

[93] SACCHERI, 2001.

[94] Ibid., p. 24.

[95] Pr. Anal. 66a 7-16: "*Consequently since the impossibility results whether the first assumption is suppressed or not, it would appear to be independent of that assumption. Or perhaps we ought not to understand the statement that the false conclusion results independently of the assumption, in the sense that if something else were supposed the impossibility would result; but rather we mean that when the first assumption is eliminated, the same impossibility results through the remaining premises; since it is not perhaps absurd that the same false result should follow from several hypotheses, e.g. that parallels meet, both on the assumption that the interior angle is greater than the exterior and on the assumption that a triangle contains more than two right angles.*"

by another similarly false; here, to exemplification purpose, he considers two possible hypotheses: the first one denies the theorem of the external angle, the second one, equivalent to the first one, suppose that the sum of the angles of a triangle is greater then two right angles; both these hypotheses are such to lead to the same contradiction: *the incidence of two straight lines already supposed parallel* ("*the parallel straight lines meet*").

7. Conclusion.

We can now wonder if the analysis until here taken allows us to answer, at least partly, to the questions that were set: that is to appraise how much of originality there is in Euclid in comparison to the preceding tradition and if his work must be considered as the summa of a scientific thought developed in the precedent century, or as an opening of a new phase of the science.
We don't say nothing new if we individuate the matter of the parallel lines as that by which the originality of the Euclidean contribution emerges with evidence, but we want to understand if he gave only a technically correct answer to a preexistent well posed problem, or instead the solution itself determined a qualitative jump on the conception of this problem, of the geometry, of the sciences altogether.
We will try to formulate an answer to different levels. A first level of novelty of the Euclidean solution, can be deduced already by the comparison with the passages of Aristotle in which a not Euclidean hypothesis is formulated. These, as we have seen, don't consider any distinction between the two connected problems of the parallelism: that is the existence and the oneness of the parallel straight line. In the *Elements*, instead, the first 28 propositions, rigorously proved without the use of the fifth postulate, seem to be placed just in that precise order to underline the impossibility, in virtue only of the first four postulates, of that we have for brevity more times pointed out as hypothesis of the obtuse angle. Rather the twenty-eighth proposition *could* implicitly constitute an answer to Aristotle's passage quoted by Toth. This hypothesis however appears not well formulated if we remember that Euclid doesn't was able to read an Aristotelian work that was published about two century later[96], so we reformulate it in the following way: Euclid had securely known the open problem that only shortly Aristotle indicated, so to this problem must be addressed the answer of Euclid that, as we said, strongly remarks the separation of the problem of the existence from that one of the oneness. Only at this point, with the twenty-ninth proposition, the fifth postulate enters the scene to answer to the second question of the parallelism.
This answer immediately calls a new question: why Euclid, certainly conscious to have given an innovative answer to the science, doesn't take the opportunity to underline his own important contribution by a clear explanation? Why he entrusts all his ideas to only the mathematical language?
Unfortunately the mathematical writings of the immediate predecessors of Euclid are not arrived until us and we therefore don't know their style, but we know the style of immediately following scientists as Archimedes and Apollonius. In such cases there are some comments or explanations, but not to make clear some philosophical options put as base of their scientific works. To suppose that the lack, in Euclid, of any explicit explanations is owed only to the chance or to a short attitude to write, is therefore totally not convincing. So much more than in the *Phenomena*, the only case in which he put to the base an explanation, this escapes from every declaration on the *being* and on the *truth* to limit himself to *save the phenomena*. The initial page can be resumed in fact, for what it results essential to us, in the following way:

[96] See note **28**.

> "*Since the stars are perceived* [as things that are moved in a certain way, according to the proofs of the optics], *then we have to set that* [is moved… in that way] [97].

The intent of "*to save the phenomena*", giving a geometric explanation of it, could not be more evident. At this point we think that it haven't any sense to ask if this was intended in a realistic way or in an instrumentalist one according to the interpretation of Duhem. Don't have sense simply because the author doesn't tell and had not perhaps any intension to declare it!
As for the *Elements*, any justification isn't given, for instance, for the postulates of the *Optic*[98], and however here also it is not difficult to discover the intent to furnish the mathematical explanation of a phenomenon (in this case that of the vision); and the fact is not secondary that such mathematical explanation is recalled then in the *Phenomena*.
Referring for the rest to what was said on the essay of Incardona, we want only to underline a point that contributes to make transparent, on the *Optic*, the character of mathematical model for a class of phenomena. That is, as we have already seen, the choice to use a discreet model and not a continuous one because such choice appears more suitable to explain (in descriptive sense) the loss of optic resolution, until to the complete disappearance of the object, when the distance increases. Therefore we are able to affirm that works as *Optic* and *Phenomena* are set as their aim "*to save the phenomena*" without assume a position on the realistic character of the scientific knowledge. Neither we see at this point some reason to believe otherwise for the *Elements*, since the objects of the geometry seem to concern the spatial form of the bodies, as from us perceived and rationally organized. It is perhaps this one the aspect that appears the most innovative if compared to what we are able to know about the preceding epistemological status of the sciences. The fact is that the not pronouncing on the realistic character of the scientific knowledge, makes these equally valid beyond the different philosophical options in a historical age of great diversifications of the ideas. This doesn't mean however a generic indifference towards the problems of the philosophy, but rather a suspension of the judgment in comparison to the matters that, in the given historical period, divided the different schools of thought in an apparently incompatible way. Such matters are just those that pertain to the status of the possible knowledge, and are expressed, in terms of theirs most radical opposition, by the dispute among Academic (Arcesilaus-Carneades) and Stoic (Zeno-Crysippus) schools.
The fact that the sciences, as they appear in the work of Euclid, are afar to ignore the philosophical problems, it is testified by the presence, to formal level, of almost all the characteristic elements of the Aristotelian deductive structure. And nevertheless the innovative characters, in comparison to the conception of the science expressed by Aristotle, appear at this point very clear: specially for what concerns the real and sure correspondence of the scientific knowledge with the reality for itself. Euclid, in other words, misses any declaration on the *being* and on the *truth* beyond the phenomena, whereas the *truth*, with the *simplicity* and not *derivability*, was for Aristotle the fundamental requisite of a proposition to be a scientific premise, just to guarantee the truth of the whole corpus of the scientific knowledge. But the formal differentiations also can't be neglected, both in the terminology (as the expression *κοιναί ἔννοιαι*) and in what the change itself of terminology can subtend[99]. So today we can say that the setup given by Euclid to the matter of the parallel straight lines doesn't could precedently to be found neither easily approved, because also respecting the Aristotelian formal deductive plant, it failed its substantial condition: the self-evidence of the premises, self-evidence that has would revealed through a

[97] EUCLIDES, Opera Omnia, Vol. 8, (Phaenomena et scripta musica), 1916.
[98] See INCARDONA.
[99] It is the case to remark the interpretation given by Russo of another terminological innovation of Euclid that, in fact, use the word *σημεῖον* to denote the geometrical point in substitution of the old term *στιγμή*. Russo riche the conclusion that such terminological change is finalized to abandon the old realistic meaning for another technical and not realistic one.

"*not further decomposability*" in more simple propositions and put substance in the oneness and absoluteness of the "truth", essential condition of the *being* as *being*.
If we accept this key of reading, it cannot surprise that the beginning of the criticism to the fifth postulate, coincides with a resumption of the Aristotelian thought and it is always addressed to try to decompose (prove) the postulate in terms of more simple propositions (or absolutely simple in the sense of not subsequently demonstrable). On this matter it is still the case to mention from the *Eudemian Ethics* a passage in which the syllogistic derivation used in the mathematical proofs is assimilated to the relationship that intervenes between cause and effect[100]: a relationship in which the symmetry is excluded. Therefore not all possible premises that allow us to prove the twenty-ninth proposition was for Aristotle equivalent: it was necessary to look for the simplest and most evident. The history of the criticism on the fifth postulate begins therefore, for what we know, in the first century B.C., with Posidonius of Rhodes (135-50 B.C.), on the base of a demand that can perhaps be that one of Aristotle, but that Euclid evidently didn't consider necessary. That the period was the same in which the lost papers of Aristotle was found and published can be casual, but it is not a coincidence that the criticism on the fifth postulate starts with Posidonius and with his pupil, or follower, Geminus. Posidonius in fact was not a scientist in the sense of Euclid or Archimedes, but a philosopher that wanted to bring back the sciences inside the philosophy[101] and, even if put himself in the Stoic area, nevertheless he contrasted Chrysippus for many aspects and tried to recover elements of the precedent Platonic-Aristotelian tradition.
Coming back to Euclid, it seems therefore justified an interpretation that shows him as a figure of scientist (in a sense that is more modern than previously admitted) intent on produce coherent and stable results, in connection with the phenomenical reality and with the τέχνη, and under shelter by the philosophical disputes. That he, for himself, would have believed that the interrelations set among the phenomena had to reflect, or at least appreciate, the reality for itself, or that contrarily he assumed a instrumentalist position, is a question that transcends the scientific discourse to concern the individual personality of a man of which we don't now nearly anything; it is something that we perhaps don't will know never why he didn't want to declare it, at least in the scientific works that arrived to us.
Today, with our knowledges, we can say that not only the suspension of the judgment on the ontological and metaphysical themes was the mean for a more general acceptance and a great efficiency of the scientific methods, but we also know that the same suspension of the judgment on the most controversial themes has also been the paid price that allowed important texts as those of Euclid, of Apollonius and of Archimedes unlike those of Zeno, or Arcesilaus or Crysippus, to cross the filter of the immediately following centuries and to survive after two thousand-three hundred years.

---

[100].Eud. Eth. 1222 b 22-30: "*And since as in other matters the first principle is a cause of the things that exist or come into existence because of it, we must think as we do in the case of demonstrations. For example, if as the angles of a triangle are together equal to two right angles the angles of a quadrilateral are necessarily equal to four right angles, that the angles of a triangle are equal to two right angles is clearly the cause of that fact; and supposing a triangle were to change, a quadrilateral would necessarily change too for example if the angles of a triangle became equal to three right angles, the angles of a quadrilateral would become equal to six right angles, or if four, eight; also if a triangle does not change but is as described, a quadrilateral too must of necessity be as described".* It is true that, as specified some line above, the assimilation to the concept of cause for "*the immovable things*" as those of the geometry, is made by Aristotle only "*for analogy*", but a substantial asymmetry remains between the more simple object (cause) and the more complex one (effect).
[101] See also note 82.

REFERENCES

AMALDI, U., *Sui concetti di retta e di piano*, in ENRIQUES, **1912**, pp. 41-108.

ARISTOTLE, *Categoriae and De interpretatione by E.M. Edghill, Analytica priora by A.J. Jenkinson ; Analytica posteriora by G.R.G. Mure ; Topica and De sophisticis elenchis by W.A. Pickard*, Oxford University press, Cambridge-London, **1971**

BALDASSARRI, M. (Editor), *La logica stoica: testi originali con introduzione e traduzione commentata*, Como, **1984**.

BERGTSON, H., *Griechische Geschichte: von den Anfängen bis in die romische Kaisezeit, München*, 1977. Ital. transl. by C. Tommasi: *L'antica Grecia dalle origini all'ellenismo*, Il Mulino, Milano, **1989**.

BOYER, C. B., *A history of Mathematics*, 1968, Italian Transl: *Storia della matematica*, ISEDI, Milano **1976**.

ARRIGHI, G., *Note sugli Elementi di Euclide*, Atti del Convegno di Studi in memoria di G. Gemignani, Modena, **1995**, pp. 87-91.

BOURBAKI, N., Élements d'histoire des mathématiques, Paris, **1960**.

BIANCHI BANDINELLI, R., *La cultura ellenistica*, Voll. 7-8, Milano, **1977**.

CANFORA, L*., Ellenismo,* Laterza, Roma – Bari, **1995**.

COHEN, M., DRABKIN, I. E., *A source book in Greek science*, McGraw-Hill, **1948**.

COLLI, G., (Editor), Aristotele: *Organon*, Adelphi, Milano, **2003.**

DE FINETTI, B., *La logica dell'incerto*, Il Saggiatore, Milano, **1989.**

DUHEM, P., *Salvare i fenomeni : saggio sulla nozione di teoria fisica da Platone a Galileo, edizione italiana a cura di Francesco Bottin* - Roma ,1986. Saggio originale: *σώζειν τὰ φαινόμενα*, Annales de Philosophien Chrétienne, VI, pp. 113-39, 277-302, 352-77, 482514, 561-92, **1908**.

DUHEM, P., *Le systeme du monde : histoire des doctrines cosmologiques de Platon a Copernic,* Paris - **1956-1973.**

EDELSTEIN, L., *Recent Trends in the Interpretation of Ancient Science*, in WINER & NOLAND (V.), **1957**, pp. 90-121.

ENRIQUES, F., (Editor), *Questioni riguardanti le Matematiche Elementari*, Bologna: Zanichelli, **1912**.

ENRIQUES, F., (Editor), *Gli Elementi di Euclide e la critica antica e moderna*, (transl.: Zapelloni M. T; Introd. and Remarques by Enriques, F.), 3 Voll., **1912-1935**.

ENRIQUES, F., *L'evoluzione delle idee geometriche nel pensiero greco*, in ENRIQUES, **1912**, Vol. II, pp. 1-40.

ENRIQUES, F., DE SANTILLANA, G., *Compendio di storia del Pensiero scientifico*, Zanichelli, Bologna, **1937**.

EUCLIDES, *Euclidis Opera Omnia,* Editors: L. Heiberg and H. Menge, Leipzig, **1883-1916**.

FLOWER, D. A., *I lidi della conoscenza: la storia dell'antica biblioteca di Alessandria*, Bardi, Roma, 2002.

FARRINGTON, B., *Greek science : its meaning for us*, Penguin books, Hardmondsworth, **1944**

FRASER, P. M., *Ptolemaic Alexandria*, 3 Voll., Oxford, **1972**.

GUARDUCCI, A., *Della Congruenza e del movimento*, in ENRIQUES, **1912**, pp. 109-142.

GEYMONAT, L., *Storia del pensiero scientifico e filosofico*, Vol. I, Milano: Garzanti, **1973**.

GUGGENHEIMER, H., *The axioms of betweeness in Euclid*, Dialectica, 31, (1-2), **1997**, pp. 187-192.

GULLINI, G., *L'ellenismo*, Jaca Book, MILANO, 1998.

HEATH, T. L., *A history of Greek mathematics,* **1**, Oxford, **1931**.

HEATH, T. L., *The Thirteen Books of Euclid's Elements* (3 Volumes), New York, **1956**.

INCARDONA, F. (Editor), Euclide: *Ottica. Immagini di una teoria della visione*, Di Renzo, **1996**

IOPPOLO, A. M., *Opinione e scienza: il dibattito tra stoici e accademici nel 3° e 2° secolo a. C.,* Napoli, **1986**.

ISNARDI PARENTI, M., *Gli Stoici, Opere e Testimonianze*, 2 v., Torino, **1994**.

ISNARDI PARENTI, M., *Lo stoicismo ellenistico,* LATERZA, BARI, **1999**.

KLINE, M., *Mathematical Thought from Ancient to Modern Times*, **1972**, Italian trasl. By Lamberti L., edited by Conte A.: Storia del pensiero matematico, Vol I, Torino: Einaudi, **1991**.

LAKATOS, I., *Proof and Refutations. The Logic of Mathematical Discovery*, Cambrige University Press, **1976**. Italian Transl. By D. Benelli: *Dimostrazioni e Confutazioni. La logica della scoperta matematica*, edited by G. Giorello, Milano, **1979**.

LAKATOS, I., MUSGRAVE. A. (Editors), *Critica e crescita della conoscenza*, Feltrinelli, Milano, **1980**.

LÉVI, C., Les Philosophies hellénistiques, PARIS, 1997, Italian transl. by A. Taglia: *Le filosofie ellenistiche*, Torino, **2002**.

LORIA G., *Le scienze esatte nell'antichità*, Milano, **1914.**

LOSEE, J., *A Historical Introduction to the Philosophy of Science*, Third Ed., Oxford University Press, **1993**. Italian transl. by D'Agostino M.: *Filosofia della scienza: un'introduzione*, Milano, **2001**.

LLOYD, G. E. R., *Methods and Problems in Greek Science*, Cambridge, 1991. Trad Ital. di F. Aronadio e E. Spinelli: *Metodi e problemi della scienza greca,* Laterza, **1993**.

PENAM, J., *Ευκλειδου Οπτικὰ καὶ Κατοπτρικὰ*, *Euclidis Optica & Catoptrica nunquam antehac græce ædita eadem latine reddita*, Parisiis apud Andream Wechelum, **1557**.

PLATONE, *Opere Complete con testo a fronte*, edizione elettronica, Laterza, 2000.

POINCARÉ, H., *La valeur de la science,* Paris,**1914**.

PROCLO DIADOCO, *Commento al 1. libro degli Elementi di Euclide*, Edited by M. Timpanaro Cardini, Pisa, **1978**.
REALE, G., (Editor), Aristotele: *Metafisica*, Bompiani, Milano, **2000**.
RUSSELL, B., *History of Western Philosophy*, London, **1961**.
RUSSO L.,. *Sulla non autenticità delle definizioni degli enti geometrici fondamentali contenute negli 'Elementi' di Euclide*, Bollettino dei classici, Accademia dei Lincei, XIII, **1992**, 25-44.
RUSSO, L., *La rivoluzione dimenticata: il pensiero scientifico greco e la scienza moderna*, Feltrinelli, Milano, **1997**.
RUSSO, L., *The definitions of fundamental geometric entities contained in book I of Euclid's Elements*, Arch. Hist. Exact. Sci., 52, No.3, **1998**, pp.195-219.
SACCHERI, G. G.; *L'euclide emendato*, Transl. and notes by G. Boccardini, Milano, **1904**.
SACCHERI, G, *Euclide liberato da ogni macchia*, Saggio introduttivo di I. Toth & E. Cattanei ; Traduzione e apparati di P. Frigerio, , **2001**
SEIDEMBERG, A., *Did Euclid's 'Elements, Book I', develop geometry axiomatically?,* Arch. Hisry Exact Sci., 14, (4), **1975**, pp. 263-295.
TANNERY, P., *Sur l'autenticité des axioms d'Euclide*, Bulletin des Sci. Math. Et Astr., 1884, p. 162.
TOTH, I., *Aristotele e i fondamenti assiomatici della geometria*, Milano, **1997**.
TREDENNICK, H. (transl.) Aristotle MA, Harvard University Press; London, **1989**.
TRUDEAU, R. J.; *The non-Euclidean revolution*, Boston [etc.] : Birkauser, 1987. Italian transl: *La rivoluzione non euclidea*, Torino, **1991**
VERONESE, G. - *Osservazioni sui principii della geometria*, Padova, **1894**.
WINER, P. P. and NOLAND, A. (Editors); *Roots of scientific thought*. *A cultural perspective*; New York, **1957**.
ZANATTA, M., (Editor), Aristotele: Organon, UTET, Torino, **1996**.
ZEUTHEN H.G., *Die geometrische Konstruktion als 'Existenzbeweis' in der antiken Geometrie,* Mathematische Annalen, 47, **1896**.

(received September 2003)

ON THE SUBGROUP LATTICE OF AN ABELIAN FINITE GROUP

Marius Tărnăuceanu
Faculty of Mathematics
"Al.I. Cuza" University of Iaşi, Romania
e-mail: mtarnauceanu@yahoo.com

The aim of this paper is to give some connections between the structure of an abelian finite group and the structure of its subgroup lattice.

1 Preliminaries

Let $(G,+)$ be an abelian group. Then the set $\mathcal{L}(G)$ of subgroups of G is a modular and complete lattice.

Moreover, we suppose that G is finite of order n. If L_n is the divisors lattice of n, then the following function is well defined:

$$\text{ord} : \mathcal{L}(G) \longrightarrow L_n, \quad \text{ord}(H) = |H|, \text{ for any } H \in \mathcal{L}(G),$$

where by $|H|$ we denote the order of the subgroup H.

2 Main results

Proposition 1. *The following conditions are equivalent:*

(i) *G is a cyclic group.*

(ii) ord *is an one-to-one function.*

(iii) ord *is a homomorphism of the semilattice* $(\mathcal{L}(G), \cap)$ *into the semilattice* $(L_n, (,))$.

(iv) ord *is a homomorphism of the semilattice* $(\mathcal{L}(G), +)$ *into the semilattice* $(L_n, [,])$.

(v) ord *is an isomorphism of lattices.*

Proof. (i) $\Longrightarrow$ (ii) Obvious.

(ii) $\Longrightarrow$ (i) For any $d \in L_n$ let M_d be the set of elements $x \in G$ having the order d. Then the family $\{M_d \mid d \in L_n\}$ is a partition of G, therefore we have:

$$n = \sum_{d \in L_n} |M_d|. \tag{1}$$

A set M_d is nonempty if and only if there exists a cyclic subgroup H_d of G having the order d. In this situation H_d is the unique subgroup of G with the order d and we have:

$$M_d = \{x \in G \mid < x >= H_d\}.$$

It results that $|M_d| = \varphi(d)$, where φ is the Euler function. Using the relation (1) and the identity

$$n = \sum_{d \in L_n} \varphi(d),$$

we obtain that $|M_d| = \varphi(d)$, for any $d \in L_n$. For $d = n$ we have $|M_d| = \varphi(n) \geq 1$, so that G contains an element of order n.

(i) $\Longrightarrow$ (iii) Let G_1, G_2 be two subgroups of G, $d_i = |G_i|$, $i = 1, 2$ and $d = |G_1 \cap G_2|$. Since $G_1 \cap G_2$ is a subgroup of G_i, $i = 1, 2$, we obtain that d/d_i, $i = 1, 2$. If d' is a divisor of d_1 and d_2, then $d' \in L_n$, so that there exists $G' \in \mathcal{L}(G)$ with $|G'| = d'$. We have $G' \subseteq G_i$, $i = 1, 2$, therefore $G' \subseteq G_1 \cap G_2$. It results that d'/d, thus $d = (d_1, d_2)$.

(iii) $\Longrightarrow$ (ii) Let G_1, G_2 be two subgroups of G such that $\mathrm{ord}(G_1) = \mathrm{ord}(G_2)$. From (iii) we obtain that $\mathrm{ord}(G_1 \cap G_2) = \mathrm{ord}(G_i)$, $i = 1, 2$, therefore we have $G_1 = G_1 \cap G_2 = G_2$.

(i) $\Longrightarrow$ (iv) Similarly with (i) $\Longrightarrow$ (iii).

(iv) $\Longrightarrow$ (ii) Similarly with (iii) $\Longrightarrow$ (ii).

(i) $\Longrightarrow$ (v) Obvious.

Next aim is to find necessary and sufficient conditions for $\mathcal{L}(G)$ in order to be a distributive lattice, respectively a complemented lattice in which every element has a unique complement.

Lemma 1. *If $\mathcal{L}(G)$ is a distributive lattice or a complemented lattice in which every element has a unique complement, then, for any $H \in \mathcal{L}(G)$, the lattice $\mathcal{L}(H)$ has the same properties.*

Proof. The first part of the assertion is obvious.

We suppose that $\mathcal{L}(G)$ is a complemented lattice in which every element has a unique complement and let H_1 be a subgroup of H. Then $H_1 \in \mathcal{L}(G)$, thus there exists a unique subgroup $\overline{H}_1 \in \mathcal{L}(G)$ such that $H_1 \oplus \overline{H}_1 = G$. It results that $H = G \cap H = H_1 \oplus (\overline{H}_1 \cap H)$. If $\widetilde{H}_1 \in \mathcal{L}(H)$ satisfies $H_1 \oplus \widetilde{H}_1 = H$ and K is the complement of H in G, then we have:

$$G = K \oplus H = (K \oplus H_1) \oplus (\overline{H}_1 \cap H)$$

and

$$G = K \oplus H = (K \oplus H_1) \oplus \widetilde{H}_1.$$

Since the subgroup $K \oplus H_1$ has a unique complement in G, it follows that $\widetilde{H}_1 = \overline{H}_1 \cap H$; hence, H_1 has a unique complement in H.

Proposition 2. *The following conditions are equivalent:*

(i) *G is a cyclic group.*

(ii) *$\mathcal{L}(G)$ is a distributive lattice.*

Proof. (i) $\Longrightarrow$ (ii) If G is a cyclic group, then, from Proposition 1, we have $\mathcal{L}(G) \simeq L_n$, thus $\mathcal{L}(G)$ is a distributive lattice.

(ii) $\Longrightarrow$ (i) From the fundamental theorem on finitely generated abelian groups there exist (uniquely determined by G) the numbers $d_1, d_2, ..., d_k \in \mathbb{N} \backslash \{0, 1\}$ satisfying $d_1 / d_2 / .../ d_k$ and

$$G \simeq \bigtimes_{i=1}^{k} \mathbb{Z}_{d_i}.$$

We shall prove that $k = 1$. If we suppose that $k \geq 2$, then let p be a prime divisor of d_1. Since d_1/d_2, we obtain that G has a subgroup isomorphic to the group $\mathbb{Z}_p \times \mathbb{Z}_p$. Using Lemma 1, it is sufficiently to verify that $\mathcal{L}(\mathbb{Z}_p \times \mathbb{Z}_p)$ is not a distributive lattice.

If $p = 2$, then $\mathcal{L}(\mathbb{Z}_p \times \mathbb{Z}_p) = \mathcal{L}(\mathbb{Z}_2 \times \mathbb{Z}_2) \simeq M_3$, which is not a distributive lattice.

If $p \geq 3$, then let H_1, H_2, H_3 be the subgroups of $\mathbb{Z}_p \times \mathbb{Z}_p$ generated by the elements $(\hat{1}, \hat{2})$, $(\hat{1}, \hat{0})$, respectively $(\hat{0}, \hat{1})$. It is easy to see that we have:

$$H_1 \cap H_2 = H_1 \cap H_3 = \{(\hat{0}, \hat{0})\}$$
$$H_1 + H_2 = H_1 + H_3 = \mathbb{Z}_p \times \mathbb{Z}_p$$
$$H_2 \neq H_3,$$

therefore $\mathcal{L}(\mathbb{Z}_p \times \mathbb{Z}_p)$ is not a distributive lattice. Hence $k = 1$, i.e. $G \simeq \mathbb{Z}_{d_1}$ is a cyclic group.

Proposition 3. *The following conditions are equivalent:*

(i) $n = |G|$ *is square–free.*

(ii) $\mathcal{L}(G)$ *is a complemented lattice in which every element has a unique complement.*

Proof. (i) $\Longrightarrow$ (ii) If n is square–free, then $\mathcal{L}(G) \simeq \mathcal{L}(\mathbb{Z}_n) \simeq L_n$. Since L_n is a complemented lattice in which every element has a unique complement (for any $d \in L_n$ there exists a unique element $\bar{d} \in L_n$, $\bar{d} = \dfrac{n}{d}$, such that $(d, \bar{d}) = 1$ and $[d, \bar{d}] = n$), $\mathcal{L}(G)$ has the same property.

(ii) $\Longrightarrow$ (i) Let G_1 be the sum of all subgroups $H \in \mathcal{L}(G)$ which are simple groups. Then there exists a unique subgroup $\overline{G}_1 \in \mathcal{L}(G)$ such that $G_1 \oplus \overline{G}_1 = G$. We shall prove that $\overline{G}_1 = \{0\}$.

If we suppose that $\overline{G}_1 \neq \{0\}$, then there exists $x \in \overline{G}_1 \backslash \{0\}$. Using the Zorn's lemma, we obtain a maximal subgroup G_2 of $\overline{G}_1$ with property that $x \notin G_2$. From Lemma 1, there exists a unique subgroup $\overline{G}_2 \in \mathcal{L}(\overline{G}_1)$ such that $G_2 \oplus \overline{G}_2 = \overline{G}_1$. It follows that $\overline{G}_2 \neq \{0\}$.

Let $G_3 \neq \{0\}$ be a subgroup of $\overline{G}_2$. Then the inclusion $G_2 \subset G_2 \oplus G_3$ implies that $x \in G_2 \oplus G_3$. From the equality $\overline{G}_1 = G_2 \oplus G_3 = G_2 \oplus \overline{G}_2$ it results that $G_3 = \overline{G}_2$, thus $\overline{G}_2$ is a simple group. We obtain $\{0\} \neq \overline{G}_2 \subset G_1 \cap \overline{G}_1$, contrary to the fact that the sum $G_1 + \overline{G}_1$ is direct.

Hence $\overline{G_1} = \{0\}$, therefore $G = G_1$. Since G is finite, there exist $H_1, H_2, ..., H_k \in \mathcal{L}(G)$ such that H_i is a simple group, for any $i = \overline{1,k}$ and $G = \bigoplus_{i=1}^{k} H_i$. For each $i \in \{1, 2, ..., k\}$ let p_i be a prime number with the property: $H_i \simeq \mathbb{Z}_{p_i}$. Using the fact that $p_i \neq p_j$ for $i \neq j$, we obtain

$$G \simeq \bigoplus_{i=1}^{k} \mathbb{Z}_{p_i} \simeq \bigtimes_{i=1}^{k} \mathbb{Z}_{p_i} \simeq \mathbb{Z}_{p_1 p_2 ... p_k},$$

i.e. $n = |G|$ is square–free.

Let Ab, Lat be the categories of abelian groups, respectively of lattices. We have a functor $\mathcal{L} : \text{Ab} \longrightarrow \text{Lat}$ given by:

a) for an abelian group G, $\mathcal{L}(G)$ is the lattice of subgroups of G;

(b) for a homomorphism of groups $f : G_1 \longrightarrow G_2$, $\mathcal{L}(f) : \mathcal{L}(G_1) \longrightarrow \mathcal{L}(G_2)$ is the homomorphism of lattices defined by $\mathcal{L}(f)(H_1) = f(H_1)$ for any $H_1 \in \mathcal{L}(G_1)$.

Remark. $\mathcal{L}$ is an exact functor.

Proposition 4. *If $f : G_1 \longrightarrow G_2$ is an epimorphism of abelian groups and*

$$L = \{H_1 \in \mathcal{L}(G_1) / \ker f \subseteq H_1\},$$

then:

(i) *L is a sublattice of $\mathcal{L}(G_1)$;*

(ii) *the function $\widetilde{f} : L \longrightarrow \mathcal{L}(G_2)$, $\widetilde{f}(H_1) = \mathcal{L}(f)(H_1) = f(H_1)$ for any $H_1 \in L$, is an isomorphism of lattices.*

Proof. (i) Obvious.

(ii) It remains to prove only that $\widetilde{f}$ is one-to-one and onto.

Let H_1, H_1' be two elements of L such that $\widetilde{f}(H_1) = \widetilde{f}(H_1')$, i.e. $f(H_1) = f(H_1')$. For any $x \in H_1$, we have $f(x) \in f(H_1) = f(H_1')$, so that there exists $y \in H_1'$ with $f(x) = f(y)$. It results that $f(x - y) = 0$, thus $x - y \in \ker f$. Since $\ker f \subseteq H_1'$, we obtain $x \in H_1'$; hence $H_1 \subseteq H_1'$. In the same way we can check the other inclusion; therefore $\widetilde{f}$ is one–to–one.

Let H_2 be a subgroup of G_2. Then $H_1 = f^{-1}(H_2) \in L$ and, using the fact that f is onto, we obtain:

$$\tilde{f}(H_1) = f(H_1) = (f \circ f^{-1})(H_2) = H_2:$$

therefore $\tilde{f}$ is onto.

Remark. The functor $\mathcal{L}$ reflects the isomorphisms, i.e. if $f : G_1 \longrightarrow G_2$ is a homomorphism of abelian groups such that $\mathcal{L}(f) : \mathcal{L}(G_1) \longrightarrow \mathcal{L}(G_2)$ is an isomorphism of lattices, then f is an isomorphism of groups.

Next aim is to study when, for two abelian groups G, G' of the same order n, the isomorphism of lattice $\mathcal{L}(G) \simeq \mathcal{L}(G')$ implies the isomorphism of groups $G \simeq G'$.

In order to solve this problem, it is necessary to minutely study the structure of the lattice $\mathcal{L}\left(\bigtimes_{i=1}^{k} \mathbb{Z}_{d_i}\right)$, where $d_i \in \mathbb{N}\backslash\{0,1\}$, $i = \overline{1,k}$ and $d_1/d_2/.../d_k$. We shall treat only the case $k = 2$, in the case $k \geq 3$ the problem remaining open.

Let $\pi : \mathbb{Z}\times\mathbb{Z} \longrightarrow \mathbb{Z}_{d_1}\times\mathbb{Z}_{d_2}$ be the function defined by $\pi(x,y) = (\bar{x}, \hat{y})$, for any $(x,y) \in \mathbb{Z}\times\mathbb{Z}$. π is an epimorphism of groups, therefore, by Proposition 4, we have the isomorphism of lattices:

$$\mathcal{L}\left(\mathbb{Z}_{d_1}\times\mathbb{Z}_{d_2}\right) \simeq L_{d_1,d_2},$$

where $L_{d_1,d_2} = \{H \in \mathcal{L}(\mathbb{Z}\times\mathbb{Z}) \mid \ker\pi = d_1\mathbb{Z}\times d_2\mathbb{Z} \subseteq H\}$. It is a simple exercise to verify that:

Lemma 2. $\mathcal{L}(\mathbb{Z}\times\mathbb{Z}) = \{H_{p,q,r} = (p,q)\mathbb{Z} + (0,r)\mathbb{Z} \mid p,q,r \in \mathbb{N},\ q < r\}$.

From the above results, we obtain that

$$\mathcal{L}(\mathbb{Z}_{d_1}\times\mathbb{Z}_{d_2}) \simeq L_{d_1,d_2} =$$

$$= \left\{H_{p,q,r} = ((p,q)\mathbb{Z} + (0,r)\mathbb{Z} \mid p,q,r \in \mathbb{N},\ q < r,\ p/d_1,\ r/\left(d_2, \frac{d_1 q}{p}\right)\right\}.$$

Remark.

1) For two elements $H_{p,q,r}, H_{p',q',r'} \in L_{d_1,d_2}$ the following conditions are equivalent:

   (i) $H_{p,q,r} = H_{p',q',r'}$.

   (ii) $p = p',\ r/(q-q', r'), r'/(q-q', r)$.

   (iii) $(p,q,r) = (p',q',r')$.

2) For two elements $H_{p,q,r}, H_{p',q',r'} \in L_{d_1,d_2}$ the following conditions are equivalent:

   (i) $H_{p,q,r} \subseteq H_{p',q',r'}$.

   (ii) $p'/p,\ r'/\left(r, q - q'\dfrac{p}{p'}\right)$.

Let A_n be the set $\{(d_1,d_2) \in (\mathbb{N}\backslash\{0,1\})^2 \mid d_1/d_2, d_1d_2 = n\}$ and $g : A_n \to \mathbb{N}^*$ be the function defined by $g(d_1,d_2) = \operatorname{card}\mathcal{L}(\mathbb{Z}_{d_1}\times\mathbb{Z}_{d_2})$ for any $(d_1,d_2) \in A_n$. If $d_1 = p_1^{\alpha_1}p_2^{\alpha_2}...p_s^{\alpha_s}$, $d_2 = p_1^{\beta_1}p_2^{\beta_2}...p_s^{\beta_s}$ are the decompositions of d_1, respectively d_2, as a product of prime factors (i.e. $p_1 =$ prime number, $\alpha_i, \beta_i \in \mathbb{N}$, $\alpha_i \leq \beta_i$, $i = \overline{1,s}$ and $p_i \neq p_j$ for $i \neq j$), then we have the following result:

Lemma 3.

$$g(d_1,d_2) = \prod_{i=1}^{s} \frac{1}{(p_i-1)^2}[(\beta_i-\alpha_i+1)p_i^{\alpha_i+2} - (\beta_i-\alpha_i-1)p_i^{\alpha_i+1} - \\ -(\alpha_i+\beta_i+3)p_i + (\alpha_i+\beta_i+1)].$$

Corollary. *We have:*

(i) $\operatorname{card}\mathcal{L}(\mathbb{Z}_2\times\mathbb{Z}_4) = 8$.

(i) $\operatorname{card}\mathcal{L}(\mathbb{Z}_2\times\mathbb{Z}_2) = 5$.

(i) $\operatorname{card}\mathcal{L}(\mathbb{Z}_p\times\mathbb{Z}_p) = p+3$, *where p is a prime number.*

Now we can prove the main result of this paper:

Proposition 5. *Let $n \geq 2$ be a natural number, s be the number of distinct prime divisors of n and G, G' be two abelian groups of order n which have (corresponding to the fundamental theorem on finitely generated abelian groups) the decompositions:*

$$G \simeq \mathbb{Z}_{d_1} \times \mathbb{Z}_{d_2},$$

respectively

$$G' \simeq \mathbb{Z}_{d'_1} \times \mathbb{Z}_{d'_2}.$$

Then, for $s \in \{1, 2\}$, the isomorphism of lattice $\mathcal{L}(G) \simeq \mathcal{L}(G')$ implies the isomorphism of groups $G \simeq G'$.

Proof. If $p_1, p_2, ..., p_s$ are the distinct prime divisors of n and $n = p_1^{h_1} p_2^{h_2} ... p_s^{h_s}$, $d_1 = p_1^{\alpha_1} p_2^{\alpha_2} ... p_s^{\alpha_s}$, $d_2 = p_1^{\beta_1} p_2^{\beta_2} ... p_s^{\beta_s}$, $d'_1 = p_1^{\alpha'_1} p_2^{\alpha'_2} ... p_s^{\alpha'_s}$, $d'_2 = p_1^{\beta'_1} p_2^{\beta'_2} ... p_s^{\beta'_s}$ are the decompositions of n, d_1, d_2, d'_1, d'_2 as a product of prime factors (where $h_i, \alpha_i, \beta_i, \alpha'_i, \beta'_i \in \mathbb{N}$, $\alpha_i \leq \beta_i$, $\alpha'_i \leq \beta'_i$, $\alpha_i + \beta_i = \alpha'_i + \beta'_i = h_i$, $i = \overline{1, s}$), then we have:

$$g(d_1, d_2) = g(d'_1, d'_2). \tag{1}$$

Since $\mathcal{L}(\mathbb{Z}_{d_1} \times \mathbb{Z}_{d_2}) \simeq \mathcal{L}(\mathbb{Z}_{d'_1} \times \mathbb{Z}_{d'_2})$, there exists an isomorphism of lattice $f : L_{d_1,d_2} \longrightarrow L_{d'_1,d'_2}$. If we consider:

(i) $p_0 = 1,\ q_0 = 0,\ r_0 = d_2$,

then we have $H_{1,0,d_2} \subseteq H_{1,0,d}$ for any $d \in \mathbb{N}$, d/d_2, therefore

$$H_{p'_0,q'_0,r'_0} \overset{\text{not}}{=\!=} f(H_{1,0,d_2}) \subseteq f(H_{1,0,d}) \overset{\text{not}}{=\!=} H_{p'_d,q'_d,r'_d} \text{ for any } d \in \mathbb{N},\ d/d_2;$$

it results r'_d / r'_0 for any $d \in \mathbb{N}$, d/d_2, thus the number of divisors of d_2 is at most the number of divisors of r'_0, so that (because r'_0/d'_2) at most the number of divisors of d'_2;

(ii) $p_0 = d_1,\ q_0 = 0,\ r_0 = 1$,

then we have $H_{d_1,0,1} \subseteq H_{e,0,1}$ for any $e \in \mathbb{N}$, e/d_1, therefore

$$H_{p''_0,q''_0,r''_0} \overset{\text{not}}{=\!=} f(H_{d_1,0,1}) \subseteq f(H_{e,0,1}) \overset{\text{not}}{=\!=} H_{p'_e,q'_e,r'_e} \text{ for any } e \in \mathbb{N},\ e/d_1;$$

it results p'_e/p''_0 for any $e \in \mathbb{N}$, e/d_1, thus the number of divisors of d_1 is at most the number of divisors of p''_0, so that (because p''_0/d'_1) at most the number of divisors of d'_1.

Starting by the isomorphism of lattice $f^{-1} : L_{d'_1,d'_2} \longrightarrow L_{d_1,d_2}$ and making a similarly reasoning, we obtain the converses of the above inequalities. Thus, for $i \in \{1,2\}$, the number of divisors of d_i is equal with the number of divisors of d'_i, i.e. the following equalities hold:

$$(2) \qquad \begin{cases} \prod_{i=1}^{s}(\alpha_i+1) = \prod_{i=1}^{s}(\alpha'_i+1) \\ \prod_{i=1}^{s}(\beta_i+1) = \prod_{i=1}^{s}(\beta'_i+1). \end{cases}$$

If $s=1$, then, from equalities (2), we obtain $\alpha_1 = \alpha'_1$ and $\beta_1 = \beta'_1$, thus:

$$G \simeq \mathbb{Z}_{d_1} \times \mathbb{Z}_{d_2} = \mathbb{Z}_{d'_1} \times \mathbb{Z}_{d'_2} \simeq G'.$$

If $s=2$, then, from the first of the equalities (2), we obtain:

$$\alpha'_1 + 1 = u(\alpha_1+1),\ \alpha'_2 + 1 = \frac{1}{u}(\alpha_2+1),$$

where $u \in Q^*$.

If we consider the function $F : \mathbb{R}_+^* \longrightarrow \mathbb{R}_+^*$ defined by

$$\begin{aligned} F(x) &= [(h_1+3-2(\alpha_1+1)x)p_1^{(\alpha_1+1)x+1} - (h_1+1-2(\alpha_1+1)x)p_1^{(\alpha_1+1)x} - \\ &- (h_1+3)p_1 + (h_1+1)]\left[\left(h_2+3-2\frac{\alpha_2+1}{x}\right)p_2^{\frac{\alpha_2+1}{x}} - \right. \\ &\left. - \left(h_2+1-2\frac{\alpha_2+1}{x}\right)p_2^{\frac{\alpha_2+1}{x}} - (h_2+3)p_2 + (h_2+1)\right] \end{aligned}$$

for any $x \in \mathbb{R}_+^*$, then the equality (1) becomes:

$$(3) \qquad F(u) = F(1).$$

We can suppose that $u \geq 1$. On the interval $[1,\infty)$ we have $F' < 0$, so that F is an one–to–one function. Therefore the equality (3) implies $u=1$. It follows that $\alpha_1 = \alpha'_1$ and $\alpha_2 = \alpha'_2$. Since $\alpha_1 + \beta_1 = \alpha'_1 + \beta'_1 = h_1$ and

$\alpha_2 + \beta_2 = \alpha_2' + \beta_2' = h_2$, it results that $\beta_1 = \beta_1'$ and $\beta_2 = \beta_2'$. Hence we have $d_1 = d_1'$ and $d_2 = d_2'$, thus:

$$G \simeq \mathbb{Z}_{d_1} \times \mathbb{Z}_{d_2} = \mathbb{Z}_{d_1'} \times \mathbb{Z}_{d_2'} \simeq G'.$$

Remark. In the case when the number s of distinct prime divisors of n is arbitrary, the equalities (1) and (2) are not sufficient to obtain $d_1 = d_1'$ and $d_2 = d_2'$. However, if the following conditions are satisfied:

(i) $h_i = 1$ for any $i \in \{1, 2, ..., s\}$ (i.e. n is square–free)
or

(ii) $h_i = 1$ for any $i \in \{1, 2, ..., s\} \backslash \{i_0\}$
(i.e. n has the form $p_1 ... p_{i_0 - 1} p_{i_0}^{h_{i_0}} p_{i_0 + 1} ... p_s$),

then it is easy to see that the conclusion of Proposition 5 holds.

References

[1] Birkhoff, G., *Lattice theory*, Amer. Math. Soc., Providence, R.I., 1967.

[2] Grätzer, G., *General lattice theory*, Academic Press, New York, 1978.

[3] Suzuki, M., *Group theory*, I, II, Springer Verlag, Berlin, 1980, 1985.

[4] Suzuki, M., *Structure of a group and the structure of its lattice of subgroups*, Springer Verlag, Berlin, 1956.

[5] Ştefănescu, M., *Introduction to group theory* (Romanian), Editura Universităţii "Al.I. Cuza", Iaşi, 1993.

Construction of k-Hyperideals by P-Hyperoperations

H. Hedayati, R. Ameri*

Department of Mathematics, Faculty of Basic Science,

University of Mazandaran, Babolsar, Iran

e-mail : {h.hedayati, ameri}@umz.ac.ir

Abstract

In this note we present a method to construction new k-hyperideals from given k-ideals of a semiring R by using of the P-hyperoperations. Then we investigate the relationship between them. In particular, we describe all k-hyperideals of the semihyperring of the nonnegative integers.

Keywords: (semi)hyperring, k-(hyper)ideal, P-hyperoperation, weak distributive [1]

1 Introduction

Hyperstructures theory was born in 1934 when Marty [12] defined hypergroups as a generalization of groups. Also Wall in 1937 defined the notion of cyclic hypergroup. This theory has been studied in the following decades and nowadays by many mathematicians. A short review of the theory of hypergroups appears in [2]. A recent books [2], [3] and [15] contain a wealth of applications. There are applications

[1]* Correspondence Author

to the following subjects: geometry, hypergraphs, binary relations, combinatorics, codes, cryptography, probability, groups, rational algebraic functions and etc. One of the several contexts which they arise is hyperring. First M. Krasner studied hyperrings, which is a triple $(R, +, .)$, where $(R, +)$ is a canonical hypergroup and $(R, .)$ is a semigroup, such that for all $a, b, c \in R$, $a(b + c) = ab + ac, (b + c)a = ba + ca$ ([10]).

The notion of k-ideals in ordinary semirings was introduced by D. R. Latore in 1965 ([11]). Also M. K. Sen and others worked on one-sided k-ideals and maximal k-ideals of semirings ([14], [16]).

The authors in [6] introduced the notion of k-hyperideals in the sense of Krasner and obtained some related results about this notion. We now follow [6] to introduce a method to construct new k-hyperideals from given k-ideals.

In section 2 of this paper, we gather all the preliminaries of (semi)hyperrings and k-(hyper)ideals which will be used in the next sections. In section 3, we represent some methods for construction semihyperrings from semirings by P-hyperoperations and then we investigate the relationship between their k-hyperideals and k-ideals. As an important result of this section, all k-hyperideals of the nonnegative integers $\mathbb{N}^*$ as a semihyperring, constructed by P-hyperoperations, are described. In section 4, we characterize the k-hyperideals of product of semihyperrings which are made by P-hyperoperations and a family of semirings.

2 Preliminaries

A map $\circ : H \times H \longrightarrow P_*(H)$ is called *hyperoperation* or *join operation*. A *hypergroupoid* is a set H with together a (binary) hyperoperation $\circ$. A hypergroupoid $(H, \circ)$, which is associative, that is $x \circ (y \circ z) = (x \circ y) \circ z, \forall x, y, z \in H$ is called a *semihypergroup* .

A *hypergroup* is a semihypergroup such that $\forall x \in H$ we have $x \circ H = H = H \circ x$, which is called *reproduction axiom* (see [2]).

Let H be a hypergroup and K be a nonempty subset of H. Then K is said to be

a *subhypergroup* of H if itself is a hypergroup under hyperoperation "$\circ$" restricted to K. Hence it is clear that a subset K of H is a subhypergroup if and only if $aK = Ka = K$, under the hyperoperation on H.

Definition 2.1. A hyperalgebra $(R, +, .)$ is called a *semihyperring* if and only if

(i) $(R, +)$ is a semihypergroup;

(ii) $(R, .)$ is a semigroup;

(iii) $\forall a, b, c \in R,\ a.(a + b) = a.b + a.c$ and $(b + c).a = b.a + c.a$.

Remark. In Definition 2.1, if we replace (iii) by

$$\forall a, b, c \in R,\ a.(a + b) \subseteq a.b + a.c \text{ and } (b + c).a \subseteq b.c + c.a,$$

we say that R is a *weak distributive* semihyperring.

A semihyperring R is called with *zero element*, if there exists an unique element $0 \in R$ such that $0 + x = x = x + 0$ and $0x = 0 = x0$ for all $x \in R$.

A semihyperring R is called *additive commutative*, if $x + y = y + x$, $\forall x, y \in R$.

A semihyperring $(R, +, .)$ is called a *hyperring* provided $(R, +)$ is a canonical hypergroup.

Definition 2.2. A hyperring $(R, +, .)$ is called

(i)*commutative* if $a.b = b.a$ for all $a, b \in R$;

(ii)*with identity*, if there exists an element, say $1 \in R$, such that $1.x = x.1 = x$ for all $x \in R$.

Let $(R, +, .)$ be a hyperring, a nonempty subset S of R is called a *subhyperring* of R if $(S, +, .)$ is itself a hyperring.

Definition 2.3. A subhyperring I of a hyperring R is said to be a (resp. *right) left hyperideal* of R provided that (resp. $x.r \in I$) $r.x \in I$ for all $r \in R$ and for all $x \in I$. We say that I is a *hyperideal* if I is both a left and right hyperideal.

Definition 2.4.[11] Let $(R, +, .)$ be a semiring. A nonempty subset I of R is called a *left k-ideal* of R, if I is a left ideal of R and for $a \in I$ and $x \in R$ we have

$$a + x \in I \text{ or } x + a \in I \implies x \in I.$$

Similarly a right k-ideal is defined. A two sided k-ideal or simply a k-ideal is both a left and right k-ideal. We denote I as k-ideal (resp. ideal) of R by $I \lhd_k R$ (resp. $I \lhd R$).

In the sequel, by R we mean a semihyperring, unless otherwise specified.

Definition 2.5.[6] Let $(R, +, .)$ be a (weak distributive) semihyperring. A nonempty subset I of R is called

(i) a *left* (resp. *right) hyperideal* of R if and only if

(a) $(I, +)$ is a semihypergroup of $(R, +)$; and

(b) $rx \in I$ (resp. $xr \in I$), for all $r \in R$ and for all $x \in I$.

(ii) a *hyperideal* of R if it is both left and right hyperideal of R. The hyperideal I of R is denoted by $I \lhd_h R$.

(iii) a *left k-hyperideal* of R, if I is a left hyperideal of R and for $a \in I$ and $x \in R$ we have

$$a + x \approx I \text{ or } x + a \approx I \implies x \in I,$$

where by $A \approx B$ we mean $A \cap B \neq \emptyset$.

(iv) Similarly a right k-hyperideal is defined. A two sided k-hyperideal or simply a k-hyperideal is both a left and right k-hyperideal. We denote I as k-hyperideal of R by $I \lhd_{k.h} R$.

3 Construction of k-hyperideals by P-hyperoperations

In this section we apply three kinds of P-hyperoperations (which were introduced for H_v-structures in [15]) to construct semihyperrings from semirings. Then we investigate the relationship between their k-hyperideals and k-ideals .

Definition 3.1. Let $(R, +, .)$ be semiring and $\emptyset \neq P \subseteq R$. We define two hyperoperations as follows

$$x \oplus_c y = \{x + t + y \mid t \in P\},$$

$$x \odot y = x.y = xy,$$

which $\oplus_c$ is called *centre P-hyperoperation.*

Proposition 3.2. Let $(R, +, .)$ be semiring and $P \subseteq R$ be a nonempty such that $PR \subseteq P$ and $RP \subseteq P$, then $(R, \oplus_c, \odot)$ is a weak distributive semihyperring.

Proof . First, we show $(R, \oplus_c)$ is a semihypergroup. For this we prove that

$$(x \oplus_c y) \oplus_c z = x \oplus_c (y \oplus_c z).$$

For $x, y, z \in R$ we have

$$\begin{aligned}
a \in (x \oplus_c y) \oplus_c z &\implies \exists a_1 \in x \oplus_c y,\ a \in a_1 \oplus_c z \\
&\implies \exists t_1, t_2 \in P,\ a = a_1 + t_1 + z,\ a_1 = x + t_2 + y \\
&\implies a = x + t_2 + y + t_1 + z \\
&\implies a = x + t_2 + b,\ b = y + t_1 + z \in y \oplus_c z \\
&\implies a \in x \oplus_c b,\ b \in y \oplus_c z \\
&\implies a \in x \oplus_c (y \oplus_c z) \\
&\implies (x \oplus_c y) \oplus_c z \subseteq x \oplus_c (y \oplus_c z).
\end{aligned}$$

Similarly, we obtain that

$$(x \oplus_c y) \oplus_c z \supseteq x \oplus_c (y \oplus_c z).$$

Clearly $(R, \odot)$ is a semigroup, since $(R, .)$ is a semigroup and $x \odot y = xy$.

We now prove weak distributivity, that is

$$\begin{aligned}
x \odot (y \oplus_c z) &\subseteq (x \odot y) \oplus_c (x \odot z) \\
&= xy \oplus_c xz.
\end{aligned}$$

For this we have

$$\begin{aligned}
a \in x \odot (y \oplus_c z) &\implies \exists a_1 \in y \oplus_c z,\ a = x \odot a_1 = xa_1 \\
&\implies \exists t \in P,\ a = xa_1,\ a_1 = y + t + z \\
&\implies a = x(y + t + z) \\
&= xy + xt + xz \in xy \oplus_c xz \quad (\ RP \subseteq P\ ) \\
&\implies x \odot (y \oplus_c z) \subseteq xy \oplus_c xz.
\end{aligned}$$

Similarly we conclude that $(y \oplus_c z) \odot x \subseteq yx \oplus_c zx$.$\square$

Definition 3.3. Let $(R, +, .)$ be a semiring and $\emptyset \neq P \subseteq R$. We define the following hyperoperations

$$x \oplus_r y = \{x + y + t \mid t \in P\}, \quad x \oplus_l y = \{t + x + y \mid t \in P\},$$

$$x \odot y = xy,$$

which $\oplus_r$ and $\oplus_l$ are called *right P-hyperoperation* and *left P-hyperoperation* respectively.

Proposition 3.4. Let $(R, +, .)$ be a semiring and $P \subseteq R$ be a nonempty such that $PR \subseteq P$ and $RP \subseteq P$ and $x + P = P + x$, for all $x \in R$. Then $(R, \oplus_r, \odot)$ and $(R, \oplus_l, \odot)$ are weak distributive semihyperrings.

Proof. First, we prove that

$$(x \oplus_r y) \oplus_r z = x \oplus_r (y \oplus_r z).$$

For this we have

$$\begin{aligned} a \in (x \oplus_r y) \oplus_r z &\implies \exists a_1 \in x \oplus_r y,\ a \in a_1 \oplus_r z \\ &\implies \exists t_1, t_2 \in P,\ a_1 = x + y + t_1,\ a = a_1 + z + t_2 \\ &\implies \exists t_1, t_2 \in P,\ a = x + y + t_1 + z + t_2 \qquad (1) \end{aligned}$$

also we have

$$\begin{aligned} b \in x \oplus_r (y \oplus_r z) &\implies \exists b_1 \in y \oplus_r z,\ b \in x \oplus_r b_1 \\ &\implies \exists w_1, w_2 \in P,\ b_1 = y + z + w_1,\ b = x + b_1 + w_2 \\ &\implies \exists w_1, w_2 \in P,\ b = x + y + z + w_1 + w_2 \qquad (2) \end{aligned}$$

From (1) we have

$$\begin{aligned} a = x + y + t_1 + z + t_2 &= x + y + z + w_1 + t_2,\ \exists w_1 \in P \qquad (z + P = P + z) \\ &\implies a \in x \oplus_r (y \oplus_r z) \qquad \text{(by (2))} \\ &\implies (x \oplus_r y) \oplus_r z \subseteq x \oplus_r (y \oplus_r z). \end{aligned}$$

Similarly we can prove that

$$(x \oplus_r y) \oplus_r z \supseteq x \oplus_r (y \oplus_r z).$$

Clearly $(R, \odot)$ is semigroup, since $(R, .)$ is a semigroup. In a similar way to the Proposition 3.2 we can prove weak distributivity. Therefore $(R, \oplus_r, \odot)$ is a weak distributive semihyperring. Analogously we can prove that $(R, \oplus_l, \odot)$ is a weak distributive semihyperring. $\square$

Remark. In Propositions 3.2 and 3.4, if we replace the conditions $RP \subseteq P$ and $PR \subseteq P$ by $rP = P = Pr$ for all $r \in R$, then $(R, \oplus_c, \odot)$ and $(R, \oplus_r, \odot)$ and $(R, \oplus_l, \odot)$ become semihyperring.

Theorem 3.5. Let $(R, +, .)$ be a semiring with zero and P be the same as Proposition 3.2 such that $0 \in P$. Then there is a one-to-one correspondence between the k-ideals of $(R, +, .)$ containing P and k-hyperideals of $(R, \oplus_c, \odot)$.

Proof. Let I be a k-ideal of $(R, +, .)$ containing P. First we prove that $I \lhd_h (R, \oplus_c, \odot)$. Suppose that $x, y \in I$, we prove $x \oplus_c y \subseteq I$. For this we have

$$\begin{aligned} z \in x \oplus_c y &\implies \exists t \in P \subseteq I,\ z = x + t + y \\ &\implies z = x + t + y \in I \quad (\text{ since } x, t, y \in I\ ) \\ &\implies x \oplus_c y \subseteq I. \end{aligned}$$

Also if $r \in R$ and $x \in I$, then $r \odot x = rx \in I$, since $I \lhd (R, +, .)$. Thus I is a hyperideal of $(R, \oplus_c, \odot)$. We now prove that $I \lhd_{k.h} (R, \oplus_c, \odot)$. For $r \in R$ and $x \in I$ we have

$$\begin{aligned} r \oplus_c x \approx I &\implies \exists z \in r \oplus_c x \approx I \\ &\implies \exists t \in P,\ z = r + t + x,\ z \in I \\ &\implies r + t + x \in I,\ t + x \in I \\ &\implies r \in I \qquad (\text{ since } I \lhd_k (R, +, .)\ ) \\ &\implies I \lhd_{k.h} (R, \oplus_c, \odot). \end{aligned}$$

Conversely, suppose that $I \lhd_{k.h} (R, \oplus_c, \odot)$. We prove that I is a k-ideal of $(R, +, .)$ containing P. For this we have

$$\begin{aligned} x, y \in I &\implies x \oplus_c y \subseteq I \qquad (\ I \lhd_h (R, \oplus_c, \odot)\ ) \\ &\implies \forall t \in P,\ x + t + y \in I \\ &\implies x + y \in I \qquad (\ 0 \in P\ )\ . \end{aligned}$$

On the other hand

$$\begin{aligned} r \in R, x \in I &\implies r \odot x \in I \qquad (\ I \lhd_h (R, \oplus_c, \odot)\ ) \\ &\implies rx \in I. \end{aligned}$$

Also we have

$$\begin{aligned} r + x \in I, x \in I &\implies r + 0 + x \in I,\ x \in I \quad (0 \in P) \\ &\implies r \oplus_c x \approx I,\ x \in I \\ &\implies r \in I \qquad\qquad (\ I \lhd_{k.h} (R, \oplus_c, \odot) \\ &\implies I \lhd_k (R, +, .). \end{aligned}$$

We have $0 \oplus_c 0 \subseteq I$, then $\{0 + t + 0 \mid t \in P\} \subseteq I$, therefore $P \subseteq I$. □

Theorem 3.6. Let $(R, +, .)$ be a semiring with zero and P be the same as Proposition 3.4 such that $0 \in P$. Then there is a one-to-one correspondence between k-ideals of $(R, +, .)$ containing P and k-hyperideals of $(\ (R, \oplus_l, \odot)\ )$ $(R, \oplus_r, \odot)$.

Proof. The proof is similar to the proof of Theorem 3.5 by some manipulation. □

Examples. (i) Let $\mathbb{N}$ be the set of natural numbers and $2\mathbb{N} = \{2, 4, 6, 8, ...\}$. Clearly $(\mathbb{N}, +, .)$ is a semiring and $2\mathbb{N}$ is a k-ideal of $(\mathbb{N}, +, .)$. Now if $P = \{4, 8, 12, 16, ...\} \subseteq 2\mathbb{N}$, then it is easy to verify that $(\mathbb{N}, \oplus_c, \odot)$ is a weak distributive semihyperring, where for all $m, n \in \mathbb{N}$ we have

$$m \oplus_c n = \{m + k + n \mid k \in P\} \ \text{ and } \ m \odot n = mn.$$

Thus $2\mathbb{N}$ is a k-hyperideal of $(\mathbb{N}, \oplus_c, \odot)$.

(ii) Let $\mathbb{N}^* = \mathbb{N} \cup \{0\}$ and $\mathbb{N}^*[x] = \{f(x) = \sum_{i=1}^{n} a_i x^i \mid a_i \in \mathbb{N}^*\}$. Clearly $(\mathbb{N}^*[x], +, .)$ is a semiring and $< x >= \{f(x) \in \mathbb{N}^*[x] \mid a_0 = 0\}$ is a k-ideal of $(\mathbb{N}^*[x], +, .)$ generated by x. Set $P =< x^m >$ for $m \in \mathbb{N}$. Obviously, $0 \in P \subseteq < x >$. Then by Propositions 3.2 and 3.5, $(\mathbb{N}^*[x], \oplus_c, \odot)$ is a weak distributive semihyperring and $< x >$ is a k-hyperideal of $(\mathbb{N}^*[x], \oplus_c, \odot)$.

In the next theorem we describe all k-hyperideals of semihyperring of the natural numbers constructed by P-hyperoperation. For this we consider the semiring $(\mathbb{N}, +, .)$ of natural numbers by usual ordinary operations.

Theorem 3.7. Let $0 \in P \subseteq \mathbb{N}^*$ and $P\mathbb{N}^* \subseteq P$ and $\mathbb{N}^*P \subseteq P$ and $P \subseteq I$. Then I is a k-hyperideal of $(\mathbb{N}^*, \oplus_c, \odot)$ if and only if there exists $a \in \mathbb{N}^*$ such that $I = \{na \mid n \in \mathbb{N}^*\}$.

Proof. By Theorem 3.5, $I \lhd_{k.h} (\mathbb{N}^*, \oplus_c, \odot)$ if and only if $I \lhd_k (\mathbb{N}^*, +, .)$. Also by Proposition 4.1 [14], $I \lhd_k (\mathbb{N}^*, +, .)$ if and only if there exists $a \in \mathbb{N}^*$ such that $I = \{na \mid n \in \mathbb{N}^*\}$. $\square$

4 Product of k-hyperideals

In the sequel by $\prod_{i \in I} R_i$, we mean the *cartesian product* of the family $\{R_i\}_{i \in I}$. It means

$$\prod_{i \in I} R_i = \{(x_i)_{i \in I} \mid x_i \in R_i\}.$$

Proposition 4.1. Let $\{R_i\}_{i \in I}$ be a family of semirings and $P_i \subseteq R_i$ be nonempty such that $R_i P_i \subseteq P_i$ and $P_i R_i \subseteq P_i$, for all $i \in I$. For $(x_i)_{i \in I}, (y_i)_{i \in I} \in \prod_{i \in I} R_i$. Define

$$(x_i)_{i \in I} \oplus_c (y_i)_{i \in I} = \{(x_i + t_i + y_i)_{i \in I} \mid t_i \in P_i\},$$

$$(x_i)_{i \in I} \odot (y_i)_{i \in I} = (x_i y_i)_{i \in I}.$$

Then $(\prod_{i \in I} R_i, \oplus_c, \odot)$ is a weak distributive semihyperring .

Proof. First we show that $(\prod_{i\in I} R_i, \oplus_c)$ is a semihypergroup. For this we prove that

$$(x_i)_{i\in I} \oplus_c [(y_i)_{i\in I} \oplus_c (z_i)_{i\in I}] = [(x_i)_{i\in I} \oplus_c (y_i)_{i\in I}] \oplus_c (z_i)_{i\in I}.$$

We have $A \in (x_i)_{i\in I} \oplus_c [(y_i)_{i\in I} \oplus_c (z_i)_{i\in I}]$

$$\begin{aligned}
&\Longrightarrow \exists t_i \in P_i,\ A \in (x_i)_{i\in I} \oplus_c (y_i + t_i + z_i)_{i\in I}\\
&\Longrightarrow \exists t_i' \in P_i,\ A = (x_i + t_i' + y_i + t_i + z_i)_{i\in I}\\
&\Longrightarrow A \in (x_i + t_i' + y_i)_{i\in I} \oplus_c (z_i)_{i\in I}\\
&\Longrightarrow A \in [(x_i)_{i\in I} \oplus_c (y_i)_{i\in I}] \oplus_c (z_i)_{i\in I}\\
&\Longrightarrow (x_i)_{i\in I} \oplus_c [(y_i)_{i\in I} \oplus_c (z_i)_{i\in I}] \subseteq [(x_i)_{i\in I} \oplus_c (y_i)_{i\in I}] \oplus_c (z_i)_{i\in I}.
\end{aligned}$$

In a similar way, we can prove the reverse inclusion. Therefore, $(\prod_{i\in I} R_i, \oplus_c)$ is a semihypergroup. Clearly $(\prod_{i\in I} R_i, \odot)$ is a semigroup. It is enough we prove weak distributivity. For this we should prove that

$$(x_i)_{i\in I} \odot [(y_i)_{i\in I} \oplus_c (z_i)_{i\in I}] \subseteq (x_i y_i)_{i\in I} \oplus_c (x_i z_i)_{i\in I}.$$

We have $A \in (x_i)_{i\in I} \odot [(y_i)_{i\in I} \oplus_c (z_i)_{i\in I}]$

$$\begin{aligned}
&\Longrightarrow \exists t_i \in P_i,\ A \in (x_i)_{i\in I} \odot (y_i + t_i + z_i)_{i\in I}\\
&\Longrightarrow A = (x_i(y_i + t_i + z_i))_{i\in I}\\
&= (x_i y_i + x_i t_i + x_i z_i)_{i\in I}\\
&\in (x_i y_i)_{i\in I} \oplus_c (x_i z_i)_{i\in I} \qquad (\ R_i P_i \subseteq P_i\ ).
\end{aligned}$$

This completes the proof. □

Proposition 4.2. If $\{R_i\}_{i\in I}$ is a family of semirings and for all $i \in I$, $P_i \subseteq R_i$ is nonempty such that $R_i P_i \subseteq P_i$ and $P_i R_i \subseteq P_i$ and $x_i + P_i = P_i + x_i$, for all $x_i \in R_i$, then $(\prod_{i\in I} R_i, \oplus_r, \odot)$ and $(\prod_{i\in I} R_i, \oplus_l, \odot)$ are weak distributive semihyperring where

$$(x_i)_{i\in I} \oplus_r (y_i)_{i\in I} = \{(x_i + y_i + t_i)_{i\in I} \mid t_i \in P_i\},$$

$$(x_i)_{i\in I} \oplus_l (y_i)_{i\in I} = \{(t_i + x_i + y_i)_{i\in I} \mid t_i \in P_i\},$$

$$(x_i)_{i\in I} \odot (y_i)_{i\in I} = (x_i y_i)_{i\in I}.$$

Proof. First we prove that $(\prod_{i\in I} R_i, \oplus_r)$ is a semihypergroup. For this we prove that

$$(x_i)_{i\in I} \oplus_r [(y_i)_{i\in I} \oplus_r (z_i)_{i\in I}] = [(x_i)_{i\in I} \oplus_r (y_i)_{i\in I}] \oplus_r (z_i)_{i\in I}.$$

We have $A \in (x_i)_{i\in I} \oplus_r [(y_i)_{i\in I} \oplus_r (z_i)_{i\in I}]$

$$\begin{aligned}
&\Longrightarrow \exists t_i \in P_i,\ A \in (x_i)_{i\in I} \oplus_r (y_i + z_i + t_i)_{i\in I} \\
&\Longrightarrow \exists t_i' \in P_i,\ A = (x_i + y_i + z_i + t_i + t_i')_{i\in I} \\
&\Longrightarrow \exists w_i \in P_i,\ A \\
&= (x_i + y_i + w_i + z_i + t_i')_{i\in I} \quad (\text{ since } z_i + P_i = P_i + z_i\ ) \\
&\in (x_i + y_i + w_i)_{i\in I} \oplus_r (z_i)_{i\in I} \\
&\subseteq [(x_i)_{i\in I} \oplus_r (y_i)_{i\in I}] \oplus_r (z_i)_{i\in I} \\
&\Longrightarrow (x_i)_{i\in I} \oplus_r [(y_i)_{i\in I} \oplus_r (z_i)_{i\in I}] \subseteq [(x_i)_{i\in I} \oplus_r (y_i)_{i\in I}] \oplus_r (z_i)_{i\in I}.
\end{aligned}$$

Similarly, we can prove that the reverse inclusion.

Clearly $(\prod_{i\in I} R_i, \odot)$ is a semigroup. Also the weak distributivity is obtained similar to the proof of Proposition 4.1. Therefore $(\prod_{i\in I} R_i, \oplus_r, \odot)$ is a semihyperring. Analogously we can prove that $(\prod_{i\in I} R_i, \oplus_l, \odot)$ is a weak distributive semihyperring. This completes the proof. □

Remark. In Propositions 4.1 and 4.2, if we replace the conditions $R_i P_i \subseteq P_i$ and $P_i R_i \subseteq P_i$ by the condition $r_i P_i = P_i = P_i r_i$, for all $r_i \in R_i$ and for all $i \in I$, then $(\prod_{i\in I} R_i, \oplus_c, \odot)$, $(\prod_{i\in I} R_i, \oplus_r, \odot)$ and $(\prod_{i\in I} R_i, \oplus_l, \odot)$ will be semihyperrings.

Proposition 4.3. If $\{R_j\}_{j\in J}$ is a family of semirings and for all $j \in J$, $P_j \subseteq R_j$ is nonempty such that $R_j P_j \subseteq P_j$ and $P_j R_j \subseteq P_j$. Then I is a k-hyperideal of $(\prod_{j\in J} R_j, \oplus_c, \odot)$ if and only if $I = \prod_{j\in J} I_j$ such that $I_j \lhd_{k.h} (R_j, \oplus_{c_j}, \odot_j)$, where

$$x_j \oplus_{c_j} y_j = \{x_j + t_j + y_j \mid t_j \in P_j\},$$

$$x_j \odot_j y_j = x_j y_j.$$

Proof. ($\Longrightarrow$) For all $j \in J$ define

$$I_j = \{x \in R_j \mid (x_i)_{i\in J} \in I,\ \exists x_i \in R_i,\ x = x_j\}.$$

We have

$$\begin{aligned} x, y \in I &\implies \exists x_i, y_i \in R_i,\ (x_i)_{i\in J}, (y_i)_{i\in J} \in I,\ x = x_j, y = y_j \\ &\implies (x_i)_{i\in J} \oplus_c (y_i)_{i\in J} \subseteq I \qquad (I \lhd_h (\prod_{j\in J} R_j, \oplus_c, \odot)) \\ &\implies \forall t_i \in P_i,\ (x_i + t_i + y_i)_{i\in J} \in I \qquad (\forall i \in J) \\ &\implies \forall t_j \in P_j,\ x + t_j + y \in I_j \\ &\implies x \oplus_{c_j} y \subseteq I_j. \end{aligned}$$

Now suppose that

$$\begin{aligned} r_j \in R_j, x \in I_j &\implies \exists r_i \in R_i,\ (r_i)_{i\in J} \in \prod_{i\in J} R_i \text{ and } \exists x_i \in R_i,\ (x_i)_{i\in J} \in I, x = x_j \\ &\implies (r_i)_{i\in J} \odot (x_i)_{i\in J} \in I \qquad (\ I \lhd_h (\prod_{i\in J} R_i, \oplus_c, \odot)\ ) \\ &\implies (r_i x_i)_{i\in J} \in I \\ &\implies r_j x_j \in I_j \qquad (\text{ by definition of } I_j\ ). \end{aligned}$$

Therefore $I_j \lhd_h R_j$.

We now show that $I_j \lhd_{k.h} R_j$ for all $j \in J$. We have

$$\begin{aligned} r_j \in R_j,\ x_j \in I_j,\ r_j \oplus_{c_j} x_j \approx I_j &\implies \exists t_j \in P_j,\ r_j + t_j + x_j \in I_j \\ &\implies (r_j)_{j\in J} \oplus_c (x_j)_{j\in J} \approx I, \end{aligned}$$

where $(r_j)_{j\in J} \in \prod_{j\in J} R_j,\ (x_j)_{j\in J} \in \prod_{j\in J} I_j$. Then since $I \lhd_{k.h} (\prod_{j\in J} R_j, \oplus_c, \odot)$ we have

$$\begin{aligned} (r_j)_{j\in J} \in I &\implies r_j \in I_j,\ \ \forall j \in J \\ &\implies I_j \lhd_{k.h} R_j. \end{aligned}$$

($\Longleftarrow$) Suppose that $I = \prod_{j\in J} I_j$ such that $I_j \lhd_{k.h} (R_j, \oplus_{c_j}, \odot_j)$. First we prove $I \lhd_h (\prod_{j\in J} R_j, \oplus_c, \odot)$. Let $(x_j)_{j\in J}, (y_j)_{j\in J} \in I$, then

$$(x_j)_{j\in J} \oplus_c (y_j)_{j\in J} = \{(x_j + t_j + y_j)_{j\in J} \mid t_j \in P_j\} \subseteq \prod_{j\in J} I_j;$$

also we have

$$\begin{aligned} I_j \lhd_h (R_j, \oplus_{c_j}, \odot_j) &\implies \forall t_j \in P_j,\ x_j + t_j + y_j \in I_j \\ &\implies (x_j)_{j\in J} \oplus_c (y_j)_{j\in J} \subseteq I. \end{aligned}$$

Now if $(r_j)_{j\in J} \in \prod_{j\in J} R_j$ and $(x_j)_{j\in J} \in I$, then $(r_j)_{j\in J} \odot (x_j)_{j\in J} = (r_j x_j)_{j\in J} \in \prod_{j\in J} I_j$, since $r_j x_j \in I_j$ by hypothesis. We now prove that $I \lhd_{k.h} (\prod_{j\in J} R_j, \oplus_c, \odot)$. For this we have

$$\begin{aligned} (r_j)_{j\in J} \in \prod_{j\in J} R_j,\ (x_j)_{j\in J} \in I,\ (r_j)_{j\in J} \oplus_c (x_1, x_2) &\approx I \\ \implies \exists t_j \in P_j,\ (r_j + t_j + x_j)_{j\in J} &\in I = \prod_{j\in J} I_j \\ \implies \exists t_j \in P_j,\ r_j + t_j + x_j &\in I_j,\ \ \forall j \in J \\ \implies r_j \oplus_{c_j} x_j &\approx I_j,\ r_j \in R_j,\ x_j \in I_j \\ \implies r_j &\in I_j \qquad (\ I_j \lhd_{k.h} (R_j, \oplus_{c_j}, \odot_j)\ ) \\ \implies (r_j)_{j\in J} &\in \prod_{j\in J} I_j. \ \square \end{aligned}$$

Proposition 4.4. Let $\{R_j\}_{j\in J}$ be a family of semirings. Suppose that $P_j \subseteq R_j$ be nonempty such that $R_j P_j \subseteq P_j$ and $P_j R_j \subseteq P_j$ and $x_j + P_j = P_j + x_j$, for all $x_j \in R_j$ and for all $j \in J$. Then I is a k-hyperideal of $(\prod_{j\in J} R_j, \oplus_r, \odot)$ (resp. $(\prod_{j\in J} R_j, \oplus_l, \odot)$) if and only if $I = \prod_{j\in J} I_j$ such that for all $j \in J$, $I_j \lhd_{k.h} (R_j, \oplus_{r_j}, \odot_j)$, (resp. $I_j \lhd_{k.h} (R_j, \oplus_{l_j}, \odot_j)$), where

$$x_j \oplus_{r_j} y_j = \{x_j + y_j + t_j \ \mid\ t_j \in P_j\},$$

$$x_j \oplus_{l_j} y_j = \{t_j + x_j + y_j \ \mid\ t_j \in P_j\},$$

$$x_j \odot_j y_j = x_j y_j.$$

Proof. The proof is similar to the proof of Proposition 4.3. $\square$

References

[1] R. Ameri, and M. M. Zahedi, "Hyperalgebraic System", Italian Journal of Pure and Applied Mathematics, No. 6 (1999) 21-32.

[2] P. Corsini, "Prolegomena of Hypergroup Theory", second edition Aviani editor, (1993).

[3] P. Corsini, and V. Leoreanu, "Applications of Hyperstructure Theory", Kluwer Academic Publications (2003).

[4] D. Freni, " A New Characterization of the Derived Hypergroup via Strongly Regular Equivalences ", Communication In Algebra, Vol. 30, No. 8 (2002), pp. 3977-3989

[5] H. Hedayati, and R. Ameri," Fuzzy k-hyperideals ", Int. J. Pu. Appl. Math. Sci., Vol. 2, No. 2, (to appear).

[6] H. Hedayati, and R. Ameri, " On k-Hyperideals of Semihyperrings", (to appear).

[7] S. Ioudilis, Polygroups et certaines de leurs proprietes, Bull. Greek. Math. Soc., vol. 22 (1981) 95-103.

[8] J. Jantosciak, "Transposition Hypergroups: Noncommutative Join Spaces", J. of Algebra , Vol. 187 (1997) 97-119.

[9] J. Jantosciak, Homomorphism, Equivalences and Reductions in Hypergroups, Rivista di Matematica Pure ed Applicata, No. 9(1991) 23-47.

[10] M. Krasner, "Approximation des Corps Values Complets de Caracteristque $P \neq 0$ Par Ceux de Caracteristique 0", Actes due Colloque d' Algebre Superieure C.B.R.M, Bruxelles, (1965) 12-22.

[11] D.R. Latore, "On h-ideals and k-ideals in hemirings", Pub. Math. Debrecen, 12 (1965) 219-226

[12] F. Marty, "Surnue generaliz-ation de la notion de group", 8^{iem} cou Scandinaves Stockholm, (1934) 45-49.

[13] I. G. Rosenberg, Hypergroups and Join Spaces determined by relations, Italian J. of Pure and Applied Math., N. 4, (1998), 93-101.

[14] M. K. Sen, and M. R. Adhikari, " On Maximal k-ideals in Semirings", Proceedings of the American Mathematics Society, Vol. 118, No. 3, July 1993.

[15] T. Vougiuklis, "Hyperstructures and their representations", Hardonic Press, Inc. (1994).

[16] H. J. Weinert, and M. K. Sen, and M. R. Adhikari, " One-sided k-ideals and h-ideals in Semirings", Mathematica Pannonica, 7/1 (1996), 147-162.

ON RECOGNITION OF CIPHER BIT STREAM FROM DIFFERENT SOURCES USING MAJORITY VOTING FUSION RULE

SHRI KANT*, VEENA SHARMA*, B. K. DASS**

*Scientific Analysis Group
Defence R & D Organization
Metcalfe House complex
Delhi-ll0054

**Deptt. of Mathematics
Faculty of mathematical Sciences
University of Delhi
Delhi-110054

Abstract

In the present paper, majority-voting rule has been investigated for its possible application in cryptological sciences. A novel approach is proposed to address the complex identification problem of overlapping classes. The method for representing patterns using different measurements has been discussed and the majority voting rule is used to fuse the results obtained in different measurement spaces. The proposed approach is quite natural and simple to implement in comparison with usual fusion strategies. The scheme has been implemented for three-class problem and results were tabulated and presented graphically.

Keywords

Decision fusion, Representation space, Pattern space, Expert classifiers, Majority logic, Stream ciphers and Cryptology.

Address for correspondence: Scientific Analysis Group, Defence R & D Organization, Metcalfe House complex, Delhi-ll0054
Tel No.: (011) 23813862
Email: shrikant@scientist.com, shrikant.ojha@gmail.com

1. Introduction

Identification of cipher bit streams generated from different sources is the primary step for a cryptanalyst. It requires cipher bit stream to be represented in the form of a pattern vector. In the measurement space, the analyst can take various measurements for patterns. Based on specific perception and scale, patterns are represented as points in some multidimensional feature space. The feature space is partitioned using the discriminant functions made on the basis of patterns of known classes, referred as training/learning patterns. The performance of the discriminant function is measured by categorization of independent patterns, known as test patterns, to their own partitions. A higher percentage of correct classification of the patterns in the test set indicates a better discriminator.

The fundamental goal of an analyst is to arrive at the highest probable correct classification of a given set of patterns. This objective leads to the design and development of different type of classifiers to solve a particular pattern recognition problem. Here, the accuracy in classification attained by different classifiers may be different. Also, the set of patterns correctly classified by one classifier may differ with the set of patterns correctly classified by another classifier. Thus, instead of searching for the best among the set of classifiers, it is found better to combine the decisions of individual classifiers. By applying a combination strategy to the set of classifiers such that the participating classifiers work complementary to each other, we are likely to get a classification rate better than that of a single best classifier.

Various combination strategies or decision fusion techniques have been proposed and studied by many researchers. Lam and Suen ([1]:1997), Kittler, et. al. ([2]:1998), Alkoot, et. al. ([3]:1999), Kuncheva, et. al. ([4]:2001), Chen and Cheng ([5]:2001) and Alexandre, et. al. ([6]:2001) etc. made a detailed study of different aspects of these combination strategies.

We first give a brief description of these fusion schemes. Let X be a pattern which is to be assigned to one of m possible classes $w_1, w_2, \ldots, w_m$ with the help of anyone of M individual classifiers. Each classifier approximates the *a posteriori* probability $P(w_i/X)$, i= 1, 2, … , m, that is the probability that pattern X belongs to class w_i, given that X was observed. A classifier assigns X to class w_k if

$$P(w_k / X) = \max_{i-1,\ldots,m} P(w_i / X) \qquad ---(1)$$

For convenience, let we denote the *a posteriori* probabilities computed by classifier C_j by $P_j(w_i/X)$, where j = 1, 2, …, M and i = 1, 2, …, m. It is assumed that these estimates of *a posteriori* probabilities given by individual classifiers are independent and identically distributed according to some pre-assumed distribution function.

Here, aim is to get improved estimates $P(w_i/X)$ by applying some combination rule '*f*' to the individual estimates $P_j(w_i/X)$ given by each of the M classifiers i.e.

$$P(w_i / X) = f\left(P_1(w_i / X), \ldots, P_j(w_i / X), \ldots, P_M(w_i / X)\right), i = 1, 2, \ldots, \mathrm{m} \quad ---(2)$$

Pattern X is finally allocated to class w_k according to the rule (1). Thus the rule for decision fusion becomes

$$X \in w_k \quad if \quad P(w_k / X) = \max_{i=1}^{m} \left\{ f\left(P_1(w_i / X), \ldots, P_M(w_i / X)\right) \right\}$$

Some prevalent decision fusion rules are the average rule, geometric mean rule, maximum rule, minimum rule, median rule, and majority vote rule. The theoretical and experimental comparative studies about the performance of decision fusion approaches have been carried out by Kittler, et. al. ([2]:1998), Alkoot and Kittler ([3]:1999), Chen & Cheng ([5]:2001) and Kuncheva ([7]:2002) etc., using different data sets. Sensitivity to estimation errors of these schemes under different

assumptions and different approximations has been analyzed, Kittler, et. al. ([2]:1998), Alkoot and Kittler ([3]:1999). It has been found that relative performance of various combination schemes changes under different conditions. The main emphasis has been given to comparison of the two basic schemes i.e. sum rule and product rule, Kittler, et. al. ([2]:1998), Alkoot and Kittler ([3]:1999), Alexandre, et. al. ([6]:2001). The sum rule is found easy to implement and less sensitive to errors than product rule, in most of the scenarios, Kittler, et. al. ([2]:1998). The product rule and strategies devised from it perform better when all the experts produce small errors. Further, the number of classifiers employed in fusion and number of classes in the problem also has an effect on the relative performance of different experts, Alkoot and Kittler ([3]:1999).

In general as stated earlier, these fusion rules, with an exception of majority voting rule, use the probabilities obtained by different classifiers to take the final fused decision about the class-memberships of the patterns. These probabilities given by different classifiers are called soft decisions. On the other hand, majority-voting rule works on hard decisions. That is, in majority voting rule, different classifiers first give their respective decisions about the class-memberships called the hard decisions, and then the decision taken by maximum number of classifiers is taken as the final decision. Instead of handling the probabilities, it simply works on the decisions given by different classifiers and therefore, is easiest to implement, Lee and Srihari ([8]:1993) and Lam and Suen ([1]:1997). And yet, experiments show that majority-voting rule is just as effective as other combination schemes, which are more complex in nature. Also, majority-voting scheme is found to be one of the schemes, which are relatively stable.

Keeping all these facts into mind, we have chosen majority-voting scheme for experimentation to support our approach of fusion, which is slightly different from the usual approach. Let us first formulate majority-voting scheme mathematically.

In majority voting rule, the individual *a posteriori* probabilities $P_j(w_i/X)$ are used to produce hard decisions δ_{ij} where

$$\delta_{ij} = \begin{cases} 1 & \text{if} \quad P_j(w_i / X) = \max_{k=1}^{m} P_j(w_k / X) \\ 0 & \text{otherwise} \end{cases}$$

Then we assign the pattern X to class w_k if

$$\left\{ \sum_{j=1}^{M} \delta_{kj} \right\} = \max_{i=1}^{m} \left\{ \sum_{j=1}^{M} \delta_{ij} \right\}$$

In the literature, mostly, two-class problem have been addressed with the help of decision fusion rules, although the rules can be implemented for any number of classes and any number of features representing a pattern. However, when the number of classes increases, the computational complexity also increases and the final decision may be costly for overlapping classes. We address this difficulty by proposing in **Section 2**, a simple and easy to implement approach, working on the basis of consensus of decisions taken in different representation spaces. **Section 3** presents the problem definition and a description about various representation spaces. **Section 4** contains the algorithm and **Section 5** contains details of experimentation and results. Finally, in **Section 6** we present our observations and conclusions.

2. Proposed Approach For Classification

Before discussing our approach, let us put the usual fusion approach in a form, which can be compared with proposed one. Let, there are m classes and M classifiers. As discussed before, the *a posteriori* probabilities p_{ij}, where i = 1, 2, …, M and j = 1, 2, …, m, are computed for a given pattern X to be classified in one of the pre-specified class.

Pattern X	Classes			
	w_1	w_2	…	w_m
Classifier C_1	p_{11}	p_{12}	…	p_{1m}
Classifier C_2	p_{21}	p_{22}	…	p_{2m}
:	:	:		:
:	:	:	…	:
Classifier C_M	p_{M1}	p_{M2}	…	p_{Mm}

Table 2.1

The pattern X gets its class membership in class w_i if a predefined function *f* as described in Section 1, gives optimum value for class w_i i.e.

$$f(p_{1i}, p_{2i}, - - -, p_{Mi}) > f(p_{1j}, p_{2j}, - - -, p_{Mj}), \quad \forall\, j \neq i$$

Now, instead of considering **different types of classifiers**, we propose to consider **different representations** of same set of patterns and allow a **single classifier** to take decision about class memberships. Going this way, in spite of having only one classifier, one can have different probable decisions and can apply any of the traditional fusion schemes. Further, if one have only two or very few classifiers available, then there will be more chances of having a tie instead of having a decision due to lack of majority of a single decision, specially when we are going to deal a multi-class problem. In that case, our approach presents a way to use fusion to have more authenticated decisions by considering many representations of set of patterns, according to the underlying problem.

As stated before, we have chosen majority voting rule for fusion i.e. we accept the decision obtained in majority of the representation/feature spaces using a single classifier. Let we have 'r' representation spaces to observe a pattern X in 'r' different ways. With the help of a classifier C, we wish to classify X in one of the pre-specified m classes. Let p_{ij}, i = 1, 2, …, r and j = 1, 2, …, m be the probability for X of membership in j^{th} class, while the i^{th} representation is used to present the pattern. First we convert these soft decisions into hard decisions $\Delta_{ij,}$ by allocating one class w_j to the pattern X in i^{th} representation space i.e.

$$\Delta_{ij} = \begin{cases} 1 & \text{if} \quad p_{ij} = \max_{k=1}^{m} p_{ik} \\ 0 & \text{otherwise} \end{cases}$$

Classifier C	Classes			
	w_1	**w_2**	---	**w_m**
Representation X_1	p_{11}	p_{12}	---	p_{1m}
Representation X_2	p_{21}	p_{22}	---	p_{2m}
Table 2.2(a)			---	
Representation X_r	p_{r1}	p_{r2}	---	p_{rm}

Soft Decisions: Table2.2(a)

Classifier C	Classes			
	w_1	**w_2**	---	**w_m**
Representation X_1	Δ_{11}	Δ_{12}	---	Δ_{1m}
Representation X_2	Δ_{21}	Δ_{22}	---	Δ_{2m}

Representation X_r	Δ_{r1}	Δ_{r2}	---	Δ_{rm}

Hard Decision: Table2.2(b)

From the table 2.2(b), it is clear that a pattern will get its class membership in class w_k if

$$\left\{ \sum_{i=1}^{r} \Delta_{ik} \right\} = \max_{j=1}^{m} \left\{ \sum_{i=1}^{r} \Delta_{ij} \right\}$$

3. Problem Definition and Feature Computation

In the present day communication scenario, any type of information viz. visual scenes, voice and text, is stored and communicated digitally. The authorized recipient at the other end recovers the same with precise accuracy and correctness. The adversary may intercept, record and retrieve all the plain transmission with some trial and error, using available means and technology. But, he will not be able to make any sense of it if the information is transmitted after encipherment by applying some cryptographic techniques. To experiment with the said problem, enciphered bit streams of scenes, voice and text have been generated from three independent stream ciphers respectively. The stream ciphers used are clock-controlled shift registers, Geffe generator and cascade of linear shift registers with nonlinear combiner. The details are described in Geffe ([9]:1973), Rueppel ([10]:1986), Schneier ([11]:1996), Kumar ([12]:1997) and Menezes et. al. ([13]:1997).

We consider each fixed length sample (now onwards referred as a message) of enciphered bit stream as a pattern. These patterns require their representation in pattern space as multidimensional feature vectors so that these can become suitable for further analysis. The process of feature extraction from each message to form a suitable mathematical pattern is like an art and this is improved by experimentation and practice. Next, we will describe the procedure followed by us to extract significant feature vectors from these bit streams.

3.1 Mathematical Representation

Let us denote the samples of enciphered binary streams by M^l_k, where l = 1, 2, 3 and k = 1, 2, …, N. In this representation, l =1 stands for encrypted scene, l =2 stands for encrypted voice and l =3 stands for encrypted text. The number of messages taken from each respective source is ‘N’. All messages are assumed to be of a sufficiently long length of ‘c’ bits, where $1000 \leq c \leq 5000$ bits usually. From each message M^l_k, binary pattern word (i.e. small blocks of bits) of a suitable fixed length ‘b’ are read,

where b = 5 or 7 bits etc. Now, these binary words can be read from a message in two (overlapping and non-overlapping) ways. In overlapped reading, we proceed bit by bit i.e. first pattern word starts from the first bit of the message and second pattern word starts from the second bit of the message and so on. And in non-overlapped reading, we move block by block i.e. we divide the whole message into blocks of given pattern word length and then these blocks are taken as pattern words.

One can take a pattern word of any length depending upon the prior knowledge of assignable character for a fixed group of bits. For a binary pattern word of length 'b', we have possibility of 2^b different words. If we do a certain computation on given message, for each of these 2^b possible words, then we will have 2^b computed quantities. These 2^b quantities or measures together will constitute a 2^b-dimensional feature vector. So, by varying pattern word length 'b', we will get feature vectors of different dimensions from a particular message. For example, for b=5 and b=7, 32-dimensional and 128-dimensional feature vector will be obtained respectively. In each case, we get different feature space with different components and different dimensions. Following this method, we can have different representations of a particular raw pattern.

In a message M^l_k, number of total occurrences of pattern words,'t' is given by

$$t = \begin{cases} c-(b-1) = t_c & (\text{Overlapping case}) \\ \left[\frac{c}{b}\right] = t_d & (\text{Non-overlapping case}) \end{cases} \quad ---(3)$$

In the subsequent sub-sections, we present further, the two different types of computations done to compute the feature vectors. In these subsections, we refer 'i^{th} pattern word' for binary equivalent of decimal number 'i', where $0 \le i \le 2^b - 1$. For example, if b = 5 then dimension of the vector =

2^5 =32 and the indices of the vector will vary from 0 to 31. It can be better understood with the help of table given below.

Binary Word	Equivalent Decimal	Feature component
00000	0	F[0]
00001	1	F[1]
|	|	|
00111	7	F[7]
01000	8	F[8]
|	|	|
11111	31	F[31]

3.1.1 Percentage Frequency Vector (PFV):

First, we compute the frequency vector F. The i^{th} component of the vector F, 'f_i' is the frequency of i^{th} pattern word in a particular message, where $0 \leq i \leq 2^b - 1$. So, i^{th} component of the percentage frequency vector P, 'p_i' is the percentage of the i^{th} component of the frequency vector F. Each component 'p_i' where $0 \leq i \leq 2^b - 1$, can be computed as

$$p_i = \begin{cases} \dfrac{f_i \times 100}{t_c} = p_i^c & (\text{Overlapping case}) \\ \\ \dfrac{f_i \times 100}{t_d} = p_i^d & (\text{Non - overlapping case}) \end{cases} \quad ---(4)$$

3.1.2 Average Distance Vector (ADV):

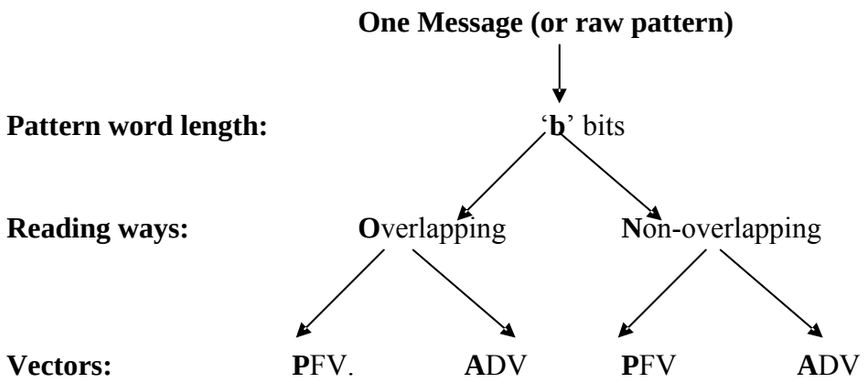
In any message, any pattern word can occur more than once. Based on these different occurrences of the same pattern word and distances between them, we compute average distance vector. Each occurrence in

the message is marked by the first bit of a pattern word. Let $P_{i,j}$ be the position of the first bit of i^{th} pattern word, in j^{th} occurrence in the message then the i^{th} component of the average distance vector **a** is defined as

$$a_i = \frac{\sum_{j=2}^{f_i} \left(P_{i,j} - P_{i,j-1} \right)}{f_i} \qquad ---(5)$$

As defined above, f_i is the number of times the i^{th} pattern word occur in the message sequence, where $0 \le i \le 2^b - 1$.

Now, from each given message bit-stream for a fixed pattern word length, four types of feature vectors can be generated with the help of (4) and (5) depending upon different choices, as shown below.



In further discussion, we will use the following notations described in table 3 to refer the four possible cases. The notations have been taken as first letter of each words near the arrow in small letters.

Notation	Pattern Word Length: b= 5 or7	Way of Reading	Type of Vector
bop	'**b**' bits	**o**verlapping	**p**ercentage frequency
boa	,,	,,	**A**verage distance
bnp	,,	**n**on-overlapping	**p**ercentage frequency
bna	,,	,,	**A**verage distance

Table 3

4. Algorithm:

Suppose there are m classes w_1, w_2, ---, w_m and N is the total number of raw patterns taken from each class. Thus, in total we have mN patterns. Let we denote the number of patterns to be taken for learning from each class by L. The remaining (N-L) patterns will be used for testing. Let we present all the patterns in 'r' different representations taking different combinations of choices i.e. varying the pattern word length, way of reading and type of vector. The dimension of each pattern in any representation will depend upon the pattern word length chosen for that representation. Let we denote the dimension in the p^{th} representation by $n_{p,}$ where p = 1, 2, ---, r.

Step 1: Make a set of raw patterns (or message bit streams), keeping the patterns of all classes together. From this set, further compute 'r' sets by converting these patterns into vectors in 'r' different representations.

Step 2: Select one of the classification technique such as minimum distance classifier, Bayes classifier or perceptron algorithm etc. as discussed in Tou and Gonzalez ([14]:1974), Bow ([15]:1984), Kant and Sharma ([16]:2000) etc.

Step 3: Pass each representation of patterns to the classifier one by one i.e. for p = 1, 2, ---, r, apply classification algorithm to p^{th}

representation which is a set containing n_p-dimensional vectors. Store class allotted to each pattern in each representation.

Step 4: Set j = 1.

Step 5: For j^{th} pattern, initialize count [i] = 0, where i = 1,2,---, m.

Step 6: Set p = 1.

Step 7: If j^{th} pattern in p^{th} representation goes to class w_k, increment the count [k] by 1.

Step 8: Repeat Step 7 for p = 2, ---, r.

Step 9: Finally, assign j^{th} pattern to class w_k if

$$\text{count [k]} = \max_{i=1}^{m} \{ \text{count [i]} \}$$

If there are more than one class such that the quantity count [k] of these classes are equal to the maximum value computed in the equation, then there arise uncertainty about the final class-membership of the pattern under consideration. In that case, the pattern is kept into the category of rejection.

Step 10: Repeat Step 5 to Step 9 for j = 2, ---, mN.

5. Experimentation and Results:

As discussed earlier, we have experimented with the problem of identification among the encrypted bit streams of scenes, speech and the text respectively. To deal with this three-class problem, we have first computed different suitable representations from these bit streams. Each representation is a set of vectors computed from the bit streams. Various techniques have been applied to classify the patterns for each

representation. Here, we are showing the classification results by two classifiers namely, maximum likelihood classifier and minimum distance classifier for each individual representation of patterns. And finally we have shown the results obtained by proposed fusion approach.

.

Maximum likelihood classifier: In the Tables 5.1(a) to 5.1(d), we have shown the percentage self-classification given by the maximum likelihood classifier for the four different representations of the same set of patterns. Notation used for each representation can be understood with the help of table 3. We have taken 150 patterns for learning of the classifier from each of the class. Table 5.2 shows the results obtained by fusion of classification results in individual representations. In Table 5.2, we have included the percentage of patterns, which cannot be allocated to any class due to uncertainty in deciding the final class membership.

% Classification	**Representation: '5na'**		
	Encrypted Scene	**Encrypted Speech**	**Encrypted Text**
Encrypted Scene	**82.67**	10	7.33
Encrypted Speech	8	**84.67**	7.33
Encrypted Text	8.67	13.33	**78**

Table 5.1(a)

% Classification	**Representation: '5oa'**		
	Encrypted Scene	**Encrypted Speech**	**Encrypted Text**
Encrypted Scene	**82.67**	9.33	10
Encrypted Speech	16.67	**80**	3.33
Encrypted Text	13.33	8.67	**78**

Table 5.1(b)

% Classification	**Representation: '7na'**		
	Encrypted Scene	**Encrypted Speech**	**Encrypted Text**
Encrypted Scene	**98**	0	2
Encrypted Speech	1	**97**	2
Encrypted Text	0.67	0	**99.33**

Table 5.1(c)

% Classification	Representation: '7oa'		
	Encrypted Scene	**Encrypted Speech**	**Encrypted Text**
Encrypted Scene	**97.33**	1.33	1.33
Encrypted Speech	1	**97**	2
Encrypted Text	0	4.67	**95.33**

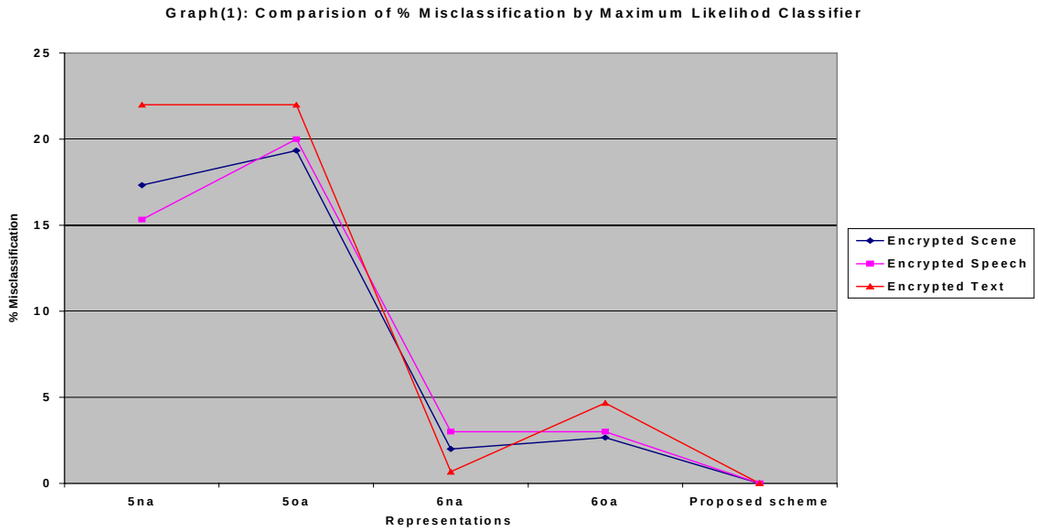
Table 5.1(d)

% Classification	Proposed Approach			
	Encrypted Scene	**Encrypted Speech**	**Encrypted Text**	**Rejected**
Encrypted Scene	**96**	0	0	4
Encrypted Speech	0	**98.67**	0	1.33
Encrypted Text	0	0	**96**	4

Table 5.2

In Tables 5.1(a) to 5.1(d), we observe some wrong classifications i.e. the percentage of patterns, which are misclassified to other classes to whom they do not belong actually. But in Table 5.2, we can see that there are no wrong classifications among classes, though we have some rejections here. This means that misclassification occurred in case of individual representations is somewhat corrected by our approach of fusion. And the patterns, which cannot be still correctly classified due to lack of consensus, are shifted to rejection category. Knowing that a misclassification is costly than a rejection, we found our classification approach to be advantageous.

This phenomenon is illustrated in the Graph(1) displayed next. In the graph, three series are plotted to show the percentage of number of misclassified patterns in each of the three classes. In each series, the classification results obtained in individual representations and by proposed fusion approach are compared. It is clear from the graph that using the proposed approach of fusion, we get a decrease to zero in percentage of misclassification in each of the class.



Minimum Distance Classifier: The percentage self-classification for different representation of patterns with minimum distance classifier has been summarized in Table 5.3(a) to 5.3(f). After fusing the classification results in these six representations, we get improved results as shown in Table 5.4. Here also, we observe that by using fusion there is a great decrement in number of misclassified patterns, in each of the class. The patterns, which cannot be allocated to any class due to a tie of votes, are kept in rejection category

% Classification	**Representation: '5oa'**		
	Encrypted Scene	**Encrypted Speech**	**Encrypted Text**
Encrypted Scene	**34**	31.33	34.67
Encrypted Speech	27.33	**48**	24.67
Encrypted Text	25.33	30.67	**44**

Table 5.3(a)

% Classification	Representation: '5na'		
	Encrypted Scene	**Encrypted Speech**	**Encrypted Text**
Encrypted Scene	**55.33**	18	26.67
Encrypted Speech	28.67	**50**	21.33
Encrypted Text	28.66	22	**49.33**

Table 5.3(b)

% Classification	Representation: '7na'		
	Encrypted Scene	**Encrypted Speech**	**Encrypted Text**
Encrypted Scene	**52.67**	24	23.33
Encrypted Speech	20	**60**	20
Encrypted Text	26	22	**52**

Table 5.3(c)

% Classification	Representation: '7na'		
	Encrypted Scene	**Encrypted Speech**	**Encrypted Text**
Encrypted Scene	**66.67**	20.67	12.67
Encrypted Speech	21.33	**55.33**	23.33
Encrypted Text	20.67	20	**59.33**

Table 5.3(d)

% Classification	Representation: '5np'		
	Encrypted Scene	**Encrypted Speech**	**Encrypted Text**
Encrypted Scene	**51.33**	25.33	23.33
Encrypted Speech	24.67	**47.33**	28
Encrypted Text	28	26.67	**45.33**

Table 5.3(e)

% Classification	Representation: '7np'		
	Encrypted Scene	**Encrypted Speech**	**Encrypted Text**
Encrypted Scene	**66.67**	16.67	16.67
Encrypted Speech	16.67	**64.67**	18.67
Encrypted Text	20	16	**64**

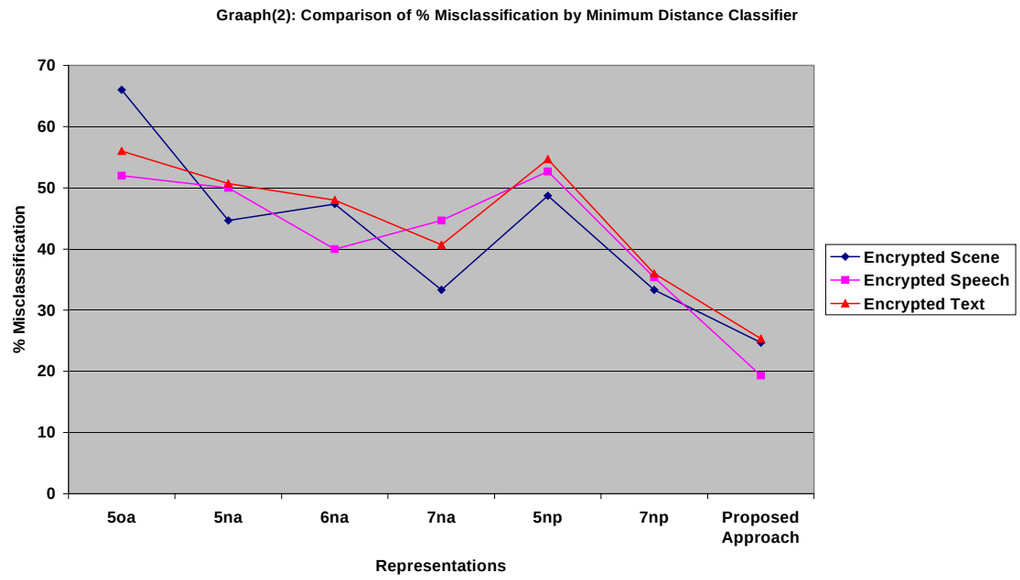
Table 5.3(f)

% Classification	Proposed Approach			
	Encrypted Scene	**Encrypted Speech**	**Encrypted Text**	**Rejected**
Encrypted Scene	**62.67**	12.67	12	12.67
Encrypted Speech	8	**60**	11.33	20.67
Encrypted Text	14.67	10.67	**59.33**	15.33

Table 5.4

With minimum distance classifier, we are able to get more than 55% classification consistently for each class. These results of classification may not be very high for the practical application, but this consistency is extremely useful from a cryptanalysis point of view.

The Graph(2) plotted to compare the results by minimum distance classifier, in different representations and those obtained by fusion, presents a similar trend as shown by maximum likelihood classifier. As compared to individual representations, the proposed fusion approach gives the least number of wrongly classified patterns.



Using both the classifiers, we have tested several sets of patterns from each class, and the test-classification is also found to be quite encouraging.

6. Observations and Conclusion:

The experimentation done for the present work has given us enough idea about handling the problem of discrimination among various random sources. The proposed idea is quite general in nature and can be applied to other kind of classification problem as well, if it is possible to compute different measurements for the same set of patterns. Also, experimentation can be done for any number of classes as described in Algorithm, instead of restricting to a three class problem. According to the nature of underlying problem and knowledge of significant features, different measurements may be computed to get different representations of

patterns. For our problem, we adhered to the most suitable representations of patterns where the classification is more transparent. Dealing with the said problem, the following observations and constraints are found to be important:

1. While assigning class membership to a pattern in each of the representation space, the classifier has three possibilities. Either the pattern will be correctly classified, wrongly classified or the classifier will remain uncertain about the class membership of the pattern. The third possibility of uncertainty of decision arises due to tie between values of discriminating function for the possible classes. This situation of neutral position of the classifier leads to no decision or rejection, i.e. classification is neither correct, nor wrong. Here, for convenience, we have considered only those representation spaces in which classifier had only two alternatives, of being correct or wrong and no rejections. Again, while taking the final decision by fusion as proposed, there may be cases of no consensus. This situation of uncertainty in deciding final class membership of a pattern leads to a rejection. We have kept these patterns in a separate category.

2. After applying proposed fusion, it has been observed that there are no wrong classifications with maximum likelihood classifier, though there are few patterns, which cannot be allocated to any class and have been kept in no decision category. Minimum distance classifier shows the similar trend with less wrong classifications by fusion as compared to those obtained in individual representations. Here also, the patterns about which the classifier is not certain are kept in rejected category. For both the classifiers, graphs have also been plotted to compare the percentage misclassification of patterns, in individual representations and after fusion by proposed approach. It is clear from the graphs by using the proposed fusion approach that we are getting reduced percentage of misclassification, which is the merit of our approach.

3. Final results, obtained by fusion by proposed approach, are better than the results obtained by using single representation spaces.

4. As we have discussed earlier, each of the representation space has dimension as n_p, p = 1, 2, ---, r. The general observation is that we obtain consistently better performance when the size of learning set is more than 5x n_p.

Acknowledgement

We would like to express our sincere gratitude and deep veneration to DR. P K Saxena**,** Director SAG and Dr. Laxmi Narain Sc. 'F' for giving us this opportunity to carry out the present work. We are also thankful to Ms. Neelam Verma, Sc. 'E' for her constructive suggestion made during the preparation this paper.

References:

[1]. Lam, L., Suen, C. Y., 1997. Application of majority voting to pattern recognition: an analysis of its behaviour and performance. IEEE transactions on Systems, Man, and Cybernetics 27(5), 553-568.

[2]. Kittler, J., Hatef, M., Duin, R., Matas, J., 1998. On combining classifiers. IEEE Trans. PAMI 20(3), 226-239.

[3]. Alkoot, F. M., Kittler, J., 1999. Experimental evaluation of expert fusion strategies. Pattern Recognition Letters 20, 1361-1369.

[4]. Kuncheva, L., Bazdek, J., Duin, R., 2001. Decision templates for multiple classifier fusion: an experimental comparison. Pattern Recognition Letters 34(2), 299-314.

[5]. Chen, D., Cheng, X., 2001. An asymptotic analysis of some expert fusion methods. Pattern Recognition Letters 22, 901-904.

[6]. Alexandre, Luis A., Campilho, Aurelio C., Kamel, M., 2001. On combining classifiers using sum and product rules. Pattern Recognition Letters 22, 1283-1289.

[7]. Kuncheva, Ludmilla I., 2002. A theoretical study on six classifier fusion strategies. IEEE Trans. PAMI 24(2), 281-286.

[8]. Lee, D. S., Srihari, S. N., 1993. Handprinted digit recognition: A comparison of algorithms. In Proc. 3rd Int. Workshop Frontiers Handwriting Recognition. Buffalo, NY, pp. 153-162.

[9]. Geffe, P. R., 1973. How to protect data with ciphers that are really hard to break. Electronics, 46(1), 99-101.

[10]. Rueppel, R. A., 1986. Analysis & design of stream ciphers. Springer-Verlag.

[11]. Schneier, B., 1996. Applied cryptography, Second Edition John Wiley & Sons, Inc.

[12]. Kumar, I. J., 1997. Cryptology: System identification and key clustering. Agean Park Press, CA, USA.

[13]. Menezes, A. J., Van Oorschot, P. C., Vanstone, S. A., 1997. Handbook applied cryptography. CRC Press, Boca Raton.

[14]. Tou, J. T., Gonzalez, R. C., 1974. Pattern recognition principles. Addison-Wesley Publishing Company.

[15]. Bow, Sing-Tze, 1984. Pattern recognition: Application to large data-set problems. Marcel Dekker, Inc., New York & Basel.

[16]. Kant, S., Sharma, V., 2000. Discrimination among various type of encrypted bitstream. International Conference on Quality Reliability and Information Technology, 21-23 Dec, New Delhi.