POZNAŃ STUDIES
IN THE PHILOSOPHY OF THE SCIENCES AND THE HUMANITIES

VOLUME 75

EDITORS

Krzysztof Brzechczyn (assistant-editor)
Jerzy Brzeziński
Michał Cichocki (assistant-editor)
Andrzej Klawiter
Piotr Kwieciński (assistant-editor)
Krzysztof Łastowski
Leszek Nowak (editor-in-chief)
Izabella Nowakowa
Katarzyna Paprzycka (Pittsburgh)
Marcin Paprzycki (Hattiesburg)
Piotr Przybysz (assistant-editor)

ADVISORY COMMITTEE

Joseph Agassi (Tel-Aviv)
Étienne Balibar (Paris)
Wolfgang Balzer (München)
Mario Bunge (Montreal)
Nancy Cartwright (London)
Robert S. Cohen (Boston)
Francesco Coniglione (Catania)
Andrzej Falkiewicz (Wrocław)
Dagfinn Føllesdal (Oslo)
Bert Hamminga (Tilburg)
Jaakko Hintikka (Boston)
Jacek J. Jadacki (Warszawa)
Jerzy Kmita (Poznań)
Leon Koj (Lublin)
Władysław Krajewski (Warszawa)
Theo A.F. Kuipers (Groningen)
Witold Marciszewski (Warszawa)
Ilkka Niiniluoto (Helsinki)
Günter Patzig (Göttingen)
Jerzy Perzanowski (Toruń)
Marian Przełęcki (Warszawa)
Jan Such (Poznań)
Max Urchs (Konstanz)
Jan Woleński (Kraków)
Ryszard Wójcicki (Warszawa)
Georg H. von Wright (Helsinki)

This book series is partly sponsored by Adam Mickiewicz University, Poznań

The address: prof. L. Nowak, Cybulskiego 13, 60-247 Poznań, Poland.
Fax: (061) 8477-079 or (061) 8471-555
E-mail: epistemo@main.amu.edu.pl

STRUCTURALIST KNOWLEDGE REPRESENTATION

PARADIGMATIC EXAMPLES

Edited by

Wolfgang Balzer
Joseph D. Sneed
C. Ulises Moulines

Amsterdam – Atlanta, GA 2000

The paper on which this book is printed meets the requirements of "ISO 9706:1994, Information and documentation - Paper for documents - Requirements for permanence".

ISSN 0303-8157
ISBN: 90-420-0680-3 (bound)
©Editions Rodopi B.V., Amsterdam - Atlanta, GA 2000
Printed in The Netherlands

CONTENTS

Robinson, J. M. (1983). Emotion, Judgment, and Desire. *Journal of Philosophy*, 80, 731-740.
Russell, J. A. (1978). Evidence of Convergent Validity on the Dimensions of Affect. *Journal of Personality and Social Psychology*, 36, 1152-1168.
Russell, J. A. (1980). A Circumplex Model of Affect. *Journal of Personality and Social Psychology*, 39, 1161-1178.
Russell, J. A. (1989). Measures of Emotion. In: Plutchik, R., Kellerman, H. (eds.), *Emotion: Theory, Research and Experience* (Vol. 4, pp. 83-111), New York: Academic.
Russell, J. A., Mehrabian, A. (1977). Evidence for a Three-Factor Theory of Emotions. *Journal of Research in Personality*, 11, 273-294.
Sander, F. (1937). Zur neueren Gefühlslehre. *Bericht über den 15. Kongreß der Deutschen Gesellschaft für Psychologie 1936*, 23-52.
Schwarz, N. (1990). Feelings as Information: Informational and Motivational Functions of Affective States. In: Higgins, E.T., Sorrentino, R.M. (eds.), *Handbook of Motivation and Cognition* (Vol. 2, pp. 527-561), New York: Guilford.
Shoemaker, S. (1975). Functionalism and Qualia. *Philosophical Studies*, 27, 291-315.
Stegmüller, W. (1986). *Theorie und Erfahrung, Band III: Die Entwicklung des neuen Strukturalismus seit 1973*. Berlin: Springer.
Stumpf, C. (1899). Über den Begriff der Gemüthsbewegung. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, 21, 47-99.
Titchener, E. B. (1908). *Lectures on the Elementary Psychology of Feeling and Attention*. New York: McMillan.
Westmeyer, H. (1989a). Psychological Theories from a Structuralist Point of View: A First Introduction. In: Westmeyer, H. (ed.), *Psychological Theories from a Structuralist Point of View* (pp. 1-12), Berlin: Springer.
Westmeyer, H. (ed.) (1992). *The Structuralist Program in Psychology: Foundations and Applications*. Toronto: Hogrefe & Huber.
Wundt, W. (1896). *Grundriß der Psychologie*. Leipzig: Engelmann.
Wundt, W. (1900), Bemerkungen zur Theorie der Gefühle. *Philosophische Studien*, 15, 149-182.
Wundt, W. (1910). *Grundzüge der physiologischen Psychologie*. (Vol. 2, 6th edition), Leipzig: Engelmann.
Wundt, W. (1911). *Grundzüge der physiologischen Psychologie*. (Vol. 3, 6th edition), Leipzig: Engelmann.
Wundt, W. (1922). *Vorlesungen über die Menschen- und Tierseele*. (7th and 8th edition), Leipzig: Voss.

Poznan Studies in the Philosophy of the Sciences and the Humanities, 2000, Vol. 75, pp. 251-283

Pablo Lorenzano

CLASSICAL GENETICS AND THE THEORY-NET OF GENETICS[1]

ABSTRACT. This article presents a reconstruction of the so-called *classical*, *formal* or *Mendelian genetics*, which is intended to be more complete and adequate than existing reconstructions. This reconstruction has been carried out with the instruments, duly modified and extended with respect to the case under consideration, of the structuralist conception of theories. The so-called *Mendel's Laws*, as well as *linkage genetics* and *gene mapping* are formulated in a precise manner while the *global structure of genetics* is represented as a *theory-net*. In doing this the *fundamental law of genetics* is made explicit and, consequently, the claims that biology in general and genetics in particular does not contain fundamental laws are challenged. These results are of methodological, philosophical and didactical relevance.

I. Introduction

The objective of this article is to present a precise reconstruction of *classical genetics* developed fundamentally by Morgan and his collaborators.[2] Attempts at providing conceptually precise foundations of classical genetics have been made by (Woodger 1959), (Lindenmayer, Simon 1980), (Rizzotti, Zanardo 1986), (Dawe 1982), (Balzer, Dawe 1986), (Balzer, Dawe 1997), (Lorenzano 1995), and (Balzer, Lorenzano 2000). While the accounts of Woodger, Lindenmayer and Simon, and Rizzotti and Zanardo have been critizised

[1] I would like to thank Thomas Bartelborth, Bernhard Lauth, Wolfgang Balzer, and specially Theo Kuipers, Isabel Payno and C. Ulises Moulines for their helpful comments on an earlier version of this paper. Part of the material presented here draws from (Balzer, Lorenzano 2000).

[2] We consider the denomination of this theory as 'Mendelian', which is the custom and practice in textbooks on genetics, to be historically erroneous as it attributes to Gregor Mendel formulations (such as the so-called 'Mendel's Laws') and developments that were never realized by him. For an analysis of these and other questions related to the history of genetics (such as the supposed 'rediscovery' of Mendel at the beginning of the century), see the first part of (Lorenzano 1995) and the included bibliography.

elsewhere[3] the reconstructions which use the same metatheory as used herein like those of (Dawe 1982), (Balzer, Dawe 1986), and (Balzer, Dawe 1997) lack generality in not incorporing the hypotheses of multiple factors, interaction of factors and multiple alleles. This reconstruction continues, extends and emends these previous attempts, particularly that of (Balzer, Dawe 1997), with the intention of being more adequate, complete and differentiated, and more in accordance with the 'standard' formulation and application of structuralism than that of (Lorenzano 1995) and (Balzer, Lorenzano 2000).

However, in addition to the notions which appear in the 'standard exposition' of the structuralist conception of theories given in *Architectonic*, such as theory-core, intended applications, theory-element, empirical claim, specialization relation and theory-net, and whose fertility for the analysis of empirical theories has been variously demonstrated, the present paper introduces - in accordance with (Balzer, Dawe 1997) - a new relationship for the specific treatment of genetics. This relationship, baptized *refinement*, is characterized on the one hand by the *addition of further kinds of objects and* perhaps also *of further functions*, such that the 'old' objects can be defined as complex structures of 'new' objects, *and*, on the other hand, by the *introduction of further law-like assumptions* concerning the 'old' and 'new' items in addition to the 'old' laws which hold also in the 'new' models. This means that the refinement relation consists in a kind of *conceptual extension* at the same time as it consists in a *specialization* of laws.

The present article characterizes first of all the *basic theory-element of genetics*, i.e. the theory-core and the domain of intended applications of genetics, and *its empirical claim*. The set-theoretic predicate, which contains the *fundamental law of genetics*, characterizes the class of *models of genetics*, and establishes the *fundamental characteristics* of the overall genetic models. By making explicit the law of genetics, doubt is cast on the claims that biology in general (Smart 1959, 1963) and genetics in particular (Kitcher 1984) does not contain fundamental laws. Subsequently, and *through refinement* of the previous predicate, the class of models of *classical genetics* is characterized. How the *special laws of the theory may be obtained* is also shown. For the purpose of exemplification, *two lines of specialization* are characterized: the first permits the precise formulation of the so-called *Mendel's laws*, with its most important subspecializations; the second treats *linkage genetics*. In addition, a refinement of the latter is also characterized which facilitates the reconstruction of *gene mapping*. Finally, the structure of genetics is represented

[3] For an exposition and evaluation of the reconstructions of (Woodger 1959), (Lindenmayer, Simon 1980) and (Balzer, Dawe 1997), see (Lorenzano 1995). For a critical comment on the reconstruction of (Rizzotti, Zanardo 1986), see (Balzer, Lorenzano 2000).

as a *theory-net*; but although this representation is incomplete in the sense that only the effectively reconstructed refinements and specializations are presented, which pertain to *one of its branches*, namely: classical genetics, it shows that the structuralist representation format of theory-nets also applies in the domain of genetic theories and confirms the philosophical thesis that the notion of a theory-net applies to scientific theories in general. This theory-net is, on the other hand, susceptible to extension, to the extent that other refinements and specializations of genetics are to be taken into account.[4]

We use the notation proposed in (Balzer, Dawe 1997), and followed in (Lorenzano 1995) and (Balzer, Lorenzano 2000), which lends itself easily to computer implementation in a high level computer language like LISP or PROLOG.[5]

II. The Basic Theory-Element of Genetics[6]

Here we introduce the different components of the theory-core of genetics, beginning with the simplest (the class of partial potential models), and subsequently passing on to the more complicated (the class of potential models, of models, and the constraints), and we will characterize its domain of intended applications and empirical claim.[7]

[4] For instance, by means of a different refinement of the basic theory-element of genetics it is possible to characterize the class of models of *molecular genetics*. See (Balzer, Dawe 1997).

[5] (Dawe, Dawe 1994) made the first steps of such implementation by way of examples.

[6] Standard expositions of genetics may be found in (Goodenough, Levine 1974), and in (Strickberger 1985).

[7] Although from a rigorous formal point of view it would be more adequate to introduce first the class of potential models and then, through it and the theoretical–non-theoretical distinction, to define the class of partial potential models, (that is the usual procedure) we prefer here to start with the later class, in order to make explicit from the outset the 'empirical' structures (relative to that theory). On the other hand, in the structures considered we do not explicitly formulate the *auxiliary base sets* (those which have a purely mathematical interpretation such as the set of real numbers $\mathbb{R}$) and restrict ourselves to the *principal base sets* (those which obtain an empirical interpretation).

2.1 The Theory-Core of Genetics

2.1.1 The Partial Potential Models of Genetics

The class of partial potential models characterizes the point of departure for genetics. It is constituted by that which is intended to be systematized, explained and forecast.

Definition 1:
$x = \langle J, P, APP, MAT, DIST \rangle$ is a *partial potential genetics*
($x = \langle J, P, APP, MAT, DIST \rangle \in M_{pp}(G)$) if and only if

(1) J is a non-empty, finite-set ('genetic individuals': variable i)
(2) P is a non-empty, finite-set ('phenotypes': variable π)
(3) APP: $J \rightarrow P$ ('appearance': $APP(i) = \pi$)
(4) MAT: $J \times J \rightarrow$ Po(J) is a partial function ('mator': $MAT(i, i') = \langle i_1, \ldots, i_n \rangle$)
(5) $DIST$: $P \times P \rightarrow$ D(P) is a partial function ('distributor': $DIST(\pi, \pi') = \langle r_1\pi_1, \ldots, r_k\pi_k \rangle$)
(6) for all $i, i' \in J$ such that MAT is defined for $\langle i, i' \rangle$ and for all $\pi \in P$:
$DIST(APP(i), APP(i'))(\pi) = RF(\pi_i, MAT(i, i'))$
('definition of distributor in terms of relative frequencies')

Comments on this predicate:
The *objects* that occur in the predicate may be interpreted as follows:

(1) J denotes the set of *genetic individuals* that can be proper individuals as well as populations. Its distinctive characteristic is that they mate and thereby produce offspring. Under each interpretation, we can distinguish between the parental genetic individuals (be they parents or parental populations), symbolized by $PARENT_1$ and $PARENT_2$, and their offspring, symbolized by $PROGENY_1, \ldots, PROGENY_n$ (where n is the number of different genetic individuals occurring in the progeny). *PARENTS* and *PROGENY* are at this general level non-specified (either proper individuals or populations) basic objects. i is used as a variable in this set. We write, then, $i \in J$ to express that individual i occurs in the model.

(2) P denotes the set of *phenotypes*. The genetic individuals are distinguished by their appearance. The intended applications of genetics are normally limited to just a few characters; it is not required that the total genetic appearance of an individual be characterized: it is enough, then, to interpret a phenotype as one or two expressions or traits of the characters

which are effectively considered in a given application. As in the case of the individuals, we here have two parental phenotypes, symbolized by $PHENOTYPE_1$ and $PHENOTYPE_2$ and n phenotypes associated with the n different offspring:
$PHENOTYPE\_OF\_PROGENY_1, \ldots, PHENOTYPE\_OF\_PROGENY_n$.
They are regarded at this general level as primitive, non-analyzed objects, which later get endowed with additional inner structure, changing their status from unanalysed, 'last' elements to more complex, defined structures. π is used as a variable in this set. We write, then, $\pi \in P$ to express that phenotype π occurs in the model.

The *functions* that occur in the predicate are interpreted as follows:

(3) *APPEARANCE* assigns to each genetic individual its phenotype, whether parental or offspring. Subsequently, the equations are of the form:
$APPEARANCE(PARENT_1) = PHENOTYPE_1$,
$APPEARANCE(PARENT_2) = PHENOTYPE_2$,
$APPEARANCE(PROGENY_i) = PHENOTYPE\_OF\_PROGENY_j$
(where $i \leq n, j \leq k$).

(4) *MATOR* represents the transition from the parents to their progeny. It assigns to any two parents their progeny:
$MATOR(PARENT_1, PARENT_2) = \langle PROGENY_1, \ldots, PROGENY_n \rangle$,
where the number n may vary with the parents.
It is a function of pairs of objects (*PARENTS*) into a set of [illegible] (the set of $PROGENY_i$). Po(J) symbolizes the power set of J. That *MATOR* is a partial function means that it needs not be defined for all the possible arguments (i.e. for pairs $\langle i, i' \rangle$ which do not mate).

(5) *DISTRIBUTOR* describes the transition of parental phenotypes to distributions of phenotypes in the offspring. That it is a partial function means that it need not be defined for pairs $\langle \pi, \pi' \rangle$ which correspond to genetic individuals that do not mate. It is a function that maps two parental phenotypes into a genetic distribution.

Quantitative consideration of the distribution of phenotypes in progeny constitutes the point of departure in genetics. The distributions of the

phenotypes are given by means of relative frequencies.[8] The total number n of the progeny is counted in addition to the number m_i of individuals of this phenotype in the offspring. As $r_i = m_i/n$ is the relative frequency of the occurrence of that particular phenotype. To calculate a distribution of phenotypes, all the relative frequencies are collected. A distribution formally is a function which assigns each component of a given n-tuple a real number which indicates that element's "weight" or "probability of occurrence". This notion is more narrow than that of a probability distribution; for this reason we speak of *genetic* distributions, or Γ-distributions.[9] The set of all Γ-distributions over some set X is denoted by $D(X)$. $\langle \pi_1,..., \pi_k \rangle$ denotes the sequence of the phenotypes in that order. Always, if one wishes, a distribution may be written in an explicit manner as a k-tuple of numbers $\langle r_1,..., r_k \rangle$, $r_i \geq 0$, $\sum r_i = 1$, where each number r_i is the weight or probability of the phenotype number i that occurs in the corresponding sequence of the phenotypes. The notation $\langle r_1\pi_1,...,r_k\pi_k \rangle$ represents a genetic distribution in explicit form.

(6) *DISTRIBUTOR* can be defined through *MATOR* and *APPEARANCE*, and thus is not a real primitive. The distribution of phenotypes is determined as follows. Beginning with two parents $PARENT_1$, $PARENT_2$ we look at the value of $MATOR(PARENT_1, PARENT_2)$, i.e. at the set $\{PROGENY_1,..., PROGENY_n\}$ of all their offspring; we determine the value of $APPEARANCE(PROGENY_i)$ for $i \leq n$, i.e. at the phenotypes that occur in the offspring; we count the total number of the offspring as well as the number of the offspring showing a given phenotype, and calculate the relative frequency of this phenotype. The list of all the relative frequencies obtained for the different offspring then is the desired distribution of phenotypes in the offspring of $PARENT_1$ and $PARENT_2$, that is the value of $DISTRIBUTOR(PARENT_1, PARENT_2)$. This gives a precise definition of the distribution of corresponding phenotypes, which can be evaluated in a mechanical manner for given forms of *MATOR* and *APPEARANCE*. *DISTRIBUTOR* has the following form:
$DISTRIBUTOR(PHENOTYPE_1, PHENOTYPE_2) = \langle r_1\pi_1,..., r_k\pi_k \rangle =$
$\langle r_1 PHENOTYPE\_OF\_PROGENY_1,..., r_k PHENOTYPE\_OF\_PROGENY_k \rangle$,
where all r_i are real positive numbers, such that $\sum_{1 \leq i \leq k} r_i = 1$.

[8] For a phenotype $\pi \in P$ and a set of genetic individuals $X \subseteq J$, the *relative frequency of π in X*, $RF(\pi/X)$ is defined in the following manner (see Balzer, Dawe 1997)):
If X is a set of proper individuals, then
$RF(\pi/X)$ = (the number of $i \in X$, such that $APPEARANCE(i) = \pi$) over (the number of elements of X).
If X is a set of populations, then
$RF(\pi/X)$ = (the number of elements in the sets $i \in X$, for which $APPEARANCE(i) = \pi$) over (the number of elements of elements of X).

[9] We do not take over the probabilistic concept of distribution: there is no use of the general features of σ-algebras here. The reformulation is fundamentally, however, a terminological issue (see Balzer, Dawe 1997). If X is a non-empty finite set, then by a Γ-distribution over X we mean a function $p: X \rightarrow [0, 1]$, such that $\Sigma_{x \in X} p(x) = 1$. Here, [0, 1] denotes the closed interval of reals (from 0 to 1). If the elements of X are ordered in such a manner that a list $\langle x_1,..., x_n \rangle$ includes exactly all the elements of X, we could write the function values of a Γ–distribution p over X in the same order $\langle p(x_1),..., p(x_n) \rangle = \langle \alpha_1,..., \alpha_n \rangle$.
Above, the x's were either phenotypes or genotypes. For $x_i = GENOTYPE\_OF\_PROGENY_i$, for example, we wrote, $\alpha_i GENOTYPE\_OF\_PROGENY_i$, in order to determine that α_i belongs to $GENOTYPE\_OF\_PROGENY_i$. In the 'abstract' notation used here, the $GENOTYPE\_OF\_PROGENY_i$ are 'swallowed' by the distribution and reoccur as its arguments: $p(GENOTYPE\_OF\_PROGENY_i) = \alpha_i$, so that there is no need to write them down additionally.

2.1.2 The Potential Models of Genetics

The class of potential models is constituted by structures that satisfy certain structural conditions (the improper axioms) for certain concepts (the G-theoretical as well as the G-non-theoretical),[10] and for which it makes sense to ask whether they are actual models of the theory.

Definition 2:
$x = \langle J, P, G, APP, MAT, DIST, DET, COMB \rangle$ is a *potential genetics*
($x = \langle J, P, G, APP, MAT, DIST, DET, COMB \rangle \in M_p(G)$) if and only if
(1) $\langle J, P, APP, MAT, DIST \rangle \in M_{pp}(G)$
(2) G is a non-empty, finite-set ('genotypes': variable γ)
(3) DET: $G \rightarrow P$ is surjective ('determiner': $DET(\gamma) = \pi$)
(4) $COMB$: $G \times G \rightarrow D(G)$ ('combinator': $COMB(\gamma, \gamma') = \langle \alpha_1\gamma_1,..., \alpha_s\gamma_s \rangle$)

Comments on this predicate:
The *objects* that occur in the predicate may be interpreted as follows:

(1) J and P are interpreted as indicated above.

[10] An application of some formal criterion for the theoretical–non-theoretical distinction is beyond the aim of this paper. We conjecture that of the five relations *APP*, *MAT*, *DIST*, *DET*, *COMB* the first three are G-non-theoretical while *DET* and *COMB* are G-theoretical. A more delicate question concerns the status of the set of genotypes. The criteria of theoreticity developed until now apply to functions or relations only and do not work for 'sets' or 'objects' (see Sneed 1983) for two possible ways to treat these cases). Intuitively, however, it seems natural to consider the set of genotypes (as well as the later introduced factors, alleles and genes) as G-theoretical, too, because they get their meaning in and through G.

(2) G denotes the set of *genotypes*. With this we reach the theoretical level. Here we encounter the most interesting genetic concepts, such as 'factor', 'gene', 'allele', that refer to theoretical entities which may be considered accountable for the occurrence of particular phenotypes, specific characters, traits or expressions. Here again we have two parental genotypes ($GENOTYPE_1$ and $GENOTYPE_2$) and finitely many genotypes for the progeny, such that for every phenotype there are several genotypes (at least one):

$GENOTYPE\_OF\_PROGENY_1,...,GENOTYPE\_OF\_PROGENY_s$.

Similar to the phenotypes, these are regarded at this general level as primitive, non-analyzed objects, which later get endowed with additional inner structure, changing their status from unanalysed, 'last' elements to more complex, defined structures. γ is used as a variable in this set. We write, then, $\gamma \in G$ to express that γ is a genotype in the model.

The *functions* that occur in the predicate are interpreted as follows:

APPEARANCE, *MATOR* and *DISTRIBUTOR* receive the same interpretation as indicated above.

(3) *DETERMINER* is a function that assigns phenotypes to genotypes. That it should be subjective means that it is a function from G to the set P, and that there are no elements of P that are not assigned to some members of G; it is a function from G *onto* P. The genotypes determine phenotypes, but the inverse does not occur; a given phenotype may happen to be determined by several different genotypes. For this reason, different numbers have been chosen: k for the number of phenotypes and s for the number of genotypes, k being normally smaller than, or equal to, s. We have equations in the form $DETERMINER(GENOTYPE_i) = PHENOTYPE_j$ (in which $i \leq s, j \leq k$).

(4) *COMBINATOR* represents the transition from parental genotypes to genotypes in the progeny. It assigns to any two particular parental genotypes a combination or a mixture of genotypes of progeny. As in the case of the phenotypes, a quantitative, probabilistic element is needed here, too. But instead of relative frequencies, proper probabilities are used here, inasmuch as we are speaking on a theoretical level at which things in general cannot be directly observed. The difference is, in certain respects, one between experimental probabilities and expected (theoretical) probabilities. Nonetheless, there are many applications in

which the relative frequencies of previous experiments are used as data for the estimation of expected probabilities. A distribution of genotypes could be considered as a genetic distribution, i.e. as a function that assigns numbers ('weights') to the genotypes.[11] If the finitely many genotypes are ordered in a sequence $\langle \gamma_1,...,\gamma_s \rangle$, such a function could be represented by a similar sequence $\langle \alpha_1,...,\alpha_s \rangle$, $\alpha_i \geq 0$, $\sum \alpha_i = 1$, which can be written in the form $\langle \alpha_1\gamma_1,...,\alpha_s\gamma_s \rangle$, in order to make explicit the underlying ordering of the genotypes. That the genotypes $\gamma_1,...,\gamma_s$ are expected to occur in the progeny with the probabilities $\alpha_1,...,\alpha_s$, respectively, is information that is conveyed by a distribution of the genotypes. *COMBINATOR* takes the form

$COMBINATOR(GENOTYPE_1, GENOTYPE_2) = \langle \alpha_1\gamma_1,..., \alpha_s\gamma_s \rangle =$
$\langle \alpha_1 GENOTYPE\_OF\_PROGENY_1,..., \alpha_s GENOTYPE\_OF\_PROGENY_s \rangle$,

where all the α_i are positive real numbers, such that $\sum_{1 \leq i \leq s} \alpha_i = 1$.

COMBINATOR is a kind of theoretical analogue of *DISTRIBUTOR*; it is a theoretical construct, given in each non-trivial application by a definition, which represents the particular hypothesis of how the genotypes in the considered system are transmitted. *COMBINATOR* is a kind of law or law-like connection at the level of the genotypes, and for this reason must be a full function.

2.1.3 The Actual Models of Genetics

The class of models of genetics, for its part, is a subclass of the potential models of genetics, whose structures satisfy, in addition to the improper axioms, the proper axiom, i.e. the fundamental law of fit.

Definition 3:

If $\mathbf{x} = \langle J, P, G, APP, MAT, DIST, DET, COMB \rangle$ is in $M_p(CG)$, then $\mathbf{x}$ is a *genetics* ($\mathbf{x} \in M(G)$) if and only if

(1) for all $i, i' \in J$ such that *MATOR* is defined for $\langle i, i' \rangle$ and for all $\gamma, \gamma' \in G$ such that

$DET(\gamma) = APP(i)$ and $DET(\gamma') = APP(i')$:

$$COMB(\gamma, \gamma') = DIST(DET(\gamma), DET(\gamma'))$$

Comments on this predicate:

The *objects* and the *functions* may be interpreted as indicated above.

[11] See note 8.

Here, we do not presuppose any special hypothesis on the number and type of genotypes, nor on the specific forms adopted by *COMBINATOR* and *DETERMINER*; this will be left unspecified.

If it is known how the phenotypes of the progeny are related to the two parental individuals, *DISTRIBUTOR* can be defined; normally, this knowledge is trivial. If π, π' are variables for parental phenotypes and $\pi_1,..., \pi_k$ are variables for phenotypes of offspring, the following may be written:

$DISTRIBUTOR(\pi, \pi') = \langle r_1\pi_1,..., r_k\pi_k\rangle$.

The above is 'replicated' on the theoretical level by *COMBINATOR*. If γ, γ' are variables for parental genotypes and $\gamma_1,..., \gamma_s$ are variables for genotypes of progeny, we may write: $COMBINATOR(\gamma, \gamma') = \langle \alpha_1\gamma_1,..., \alpha_s\gamma_s\rangle$.

The basic axiom of genetics (1) may be read as stating that the theoretical frequencies of genotypes as produced by *COMBINATOR* should *ideally*[12] coincide with those observed in progeny as expressed in the corresponding function value of *DISTRIBUTOR*; it states equalities of genetic distributions, namely: of the distribution of genotypes $COMB(\gamma, \gamma')$ and the corresponding distribution of phenotypes $DIST(DET(\gamma), DET(\gamma'))$.

The condition (1) of the above set-theoretic predicate claims, then:

(1) For any given parental pair, the genetic distributions of genotypes - produced by *COMBINATOR* - and of phenotypes - given by *DISTRIBUTOR* - in the progeny of this pair ideally fit with each other.

This condition binds together the most important primitives of G in an inseparable way. This purely syntactical feature distinguishes (1) as a cluster-law, in fact, *the cluster-law* of G. Moreover, this cluster-law is assumed to hold in all models of G, and therefore may be seen as the *fundamental law* of G.

This observation is in contradiction with Smart's claim (Smart 1959, 1963) that biology does not have fundamental laws at all or Kitcher's claim (Kitcher 1984) that genetics does not have a fundamental law. The clue to this contradiction is in the notion of a fundamental law.[13] What is a fundamental law of a theory? From the well known examples of physical theories we know that a fundamental law is not like Smart thinks a 'law in strict sense' (in his terminology) - i.e. a non-analytical universal statement which is supposed to be applied everywhere in space and time and which can be expressed in perfect

[12] We use the expression *ideally* in order to indicate that we do not take into account features of approximation that genetics contains like practically all empirical theories.

[13] The following reproduces the arguments given in (Balzer, Lorenzano 2000).

general terms without making use of proper names or of tacit reference to proper names - but a law accepted in the scientific community which is assumed to hold in each of the theory's applications. This is compatible with, and confirms, the picture of theory-nets of the structuralist conception. A fundamental law is valid in all applications, and therefore is characteristic for the class of models of the basic theory-element of the theory-net. The point is that geneticists up to now did not formulate such a fundamental law. In the literature of genetics no such law can be made out - so far Kitcher is right.

On the other hand, our reconstructive work suggests such a fundamental law, not on historical but on systematic grounds. We have identified the axiom of fit as a law providing a frame for various specializations of *COMBINATOR* and *DETERMINER*, and we can point to the historical fact of various such specializations having occurred. Of course, these historical 'specializations' do not have the form of processes of specialization, simply because, historically, there is no basic theory-element to be specialized. Historically, what happens is the introduction of different specific genetic laws, which are valid only in some applications. But with hindsight it turns out that these laws are specializations of a basic theory-element postulated on purely systematic grounds, and without direct historical evidence. We think that this systematic argument gives some credit to our view that (1) is, in fact, the fundamental law of G.

In order to achieve a thorough understanding of this law consider two parental individuals with phenotypes π, π', genotypes γ, γ' and the corresponding genetic distributions over phenotypes and genotypes in their progeny: $d_{ph} = \langle r_1\pi_1,..., r_k\pi_k\rangle$, $d_{ge} = \langle \alpha_1\gamma_1,..., \alpha_s\gamma_s\rangle$. Consider first the simple case in which *DETERMINER* is one-one. In this case each phenotype π_j comes from exactly one of the genotypes $\gamma_1,..., \gamma_s$. So $k = s$ and we may assume that each π_j is produced by γ_j. The natural notion of fit between the two distributions $\langle r_1\pi_1,..., r_k\pi_k\rangle$, $\langle \alpha_1\gamma_1,..., \alpha_s\gamma_s\rangle$ is this. We say that d_{ph} and d_{ge} *ideally fit* with each other if and only if, for all $j \leq s$: $r_j = \alpha_j$.

In general, the situation is not as simple as that for in general the same phenotypes may be produced by different genotypes. In these cases we have to compare the probabilities of all these genotypes with the relative frequency of the phenotype they all produce. Formally, let us introduce, for given parental genotypes γ, γ', and given index $j \leq k$ the set $C(\gamma, \gamma', j)$ of all probabilities α_i occurring in d_{ge} such that the corresponding genotype γ_i produces phenotype π_j (compare Fig. 1). Moreover, let us write $c_j = \Sigma\ \alpha_i$, $i \in C(\gamma, \gamma', j)$, for the sum of all those probabilities α_i whose corresponding genotype γ_i gives rise to the same π_j with relative frequency r_j.

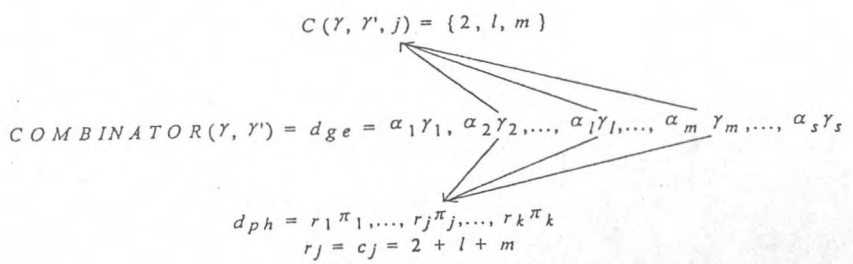

Fig. 1

In order to fit d_{ph} and d_{ge} we then have to compare each relative frequency π_j with the sum c_j. We say that, in the general case, d_{ph} and d_{ge} *ideally fit* if and only if, for all $j \leq k$: $r_j = c_j$.

Two genetic distributions of genotypes $\langle \alpha_1\gamma_1,...,\alpha_s\gamma_s \rangle$ and of phenotypes $\langle r_1\pi_1,...,r_k\pi_k \rangle$ *ideally fit* together if and only if:

i) $k \leq s$,
ii) each phenotype π_j arises from one genotype γ_i or several different genotypes γ_i by means of *DETERMINER*,
iii) the probability coefficients of the items related in ii) ideally fit with each other.

As an objection to taking (1) as a fundamental law one might point to its triviality. The axiom of fit has little empirical content. If *DISTRIBUTOR* is determined empirically and the precise form of *COMBINATOR* is hypothetically postulated the axiom of fit amounts to saying that the coefficients in the distribution of phenotypes and in that of genotypes in the offspring are equal. Checking this is a paper and pencil task and does not involve empirical work. The objection then is that such empirically empty 'laws' should be dismissed as candidates for a fundamental law, and that therefore (1) cannot be taken as the fundamental law of G. This consideration may explain why geneticists did not

consider (1) as a fundamental law, and in this way the lack of such a law in the literature of genetics.[14] But this resistance to (1) is unfounded, and results from a misconception of the role which fundamental laws play in science. Several examples from physics and economics[15] have shown that the primary role of fundamental laws is to provide a conceptual frame in which further, empirically non-trivial laws can be formulated. Fundamental laws do not primarily express empirical connections but have a rather analytical flavour. This is true for Newton's second law as well as for the assumption of utility maximization in economics. The genetic axiom of fit has exactly the same status. Connecting the theoretical concept *COMBINATOR* via *DETERMINER* with the other, empirically more accessible notions, the axiom of fit provides a frame in which further specializations concerning the form of *COMBINATOR* and *DETERMINER* can be formulated. In this respect, (1) is in good company with other, established and acknowledged fundamental laws. We take this as a second argument for our view about (1).

2.1.4 Constraints for Genetics

The *general constraint for genetics* C(G) is a kind of relationship denominated *equality-constraint*. In general, the conditions of an equality-constraint function in the following way. A function is considered that represents a property of the objects of the theory. The condition of an equality-constraint for this function requires, then, that the objects that occur in different applications possess the same value in all of these applications.

In the case considered, there is the requirement that the same phenotypes should be assigned to the same genotypes in all genetic applications in which they occur. We use here the symbol C (for 'constraint') and the following conventions: if x or y is an element of the set of potential models of G, then their corresponding components should have x or y as subindices. The set of genotypes of the potential model x is symbolized by G_x, the *DETERMINER* function of y is symbolized by $DETERMINER_y$, etc. The constraint is symbolized by C_{DET}. The subindex indicates the function to be treated, that is

[14] For instance, Thomas Hunt Morgan argued in 1909 that the Mendelian theory was nothing more than a logical construct, a conceptualization dealing with formalistic symbols which had no basis in reality. He wrote: "In the modern interpretation of Mendelism, facts are transformed into factors at a rapid rate. If one factor will not explain the facts, then two are invoked; if two proved insufficient, three will sometimes work out. The superior jugglery sometimes necessary to account for the results are often so excellently 'explained' because the explanation was invented to explain them and then presto! explain the facts by the very factors that we invented to account for them." See (Morgan 1967, p. 365).

[15] For classical particle mechanics and general equilibrium theory, see *Architectonic*.

the *DETERMINER* function; and the supraindex indicates the kind of constraint, that is: of equality. This constraint is expressed in the following manner:

Definition 4:

The *equality constraint* $C_{DET}^{(\approx,=)}$ *for determiner* is defined by

$X \in C_{DET}^{(\approx,=)}$ if and only if $X \subseteq M_p(G)$ and for all $x, y \in X$ and all γ, if $\gamma \in G_x \cap G_y$, then $DETERMINER_x(\gamma) = DETERMINER_y(\gamma)$.

In a truly complete reconstruction of G we should include the links this theory has to other (underlying) theories. However, since in this article we leave open the question of the essential links of G to other theories and make the idealizing assumption that there are no such links, the *theory-core of genetics* (K(G)) can be characterized as follows:

$$K(G): = \langle M_p(G), M(G), M_{pp}(G), C(G) \rangle.$$

2.2 The Intended Applications and the Basic Theory-Element of Genetics

The *domain of intended applications* constitutes the class of those empirical systems to which one wishes to apply the fundamental laws of the theory. They cannot be characterized by purely formal means. The only thing that we *can* say from a formal point of view is that an intended application is a partial potential model. In our case that means that $I(G) \subseteq M_{pp}(G)$ and that the members of I(G) - to which one wishes to apply the fundamental law of fit - are real systems containing genetic individuals (proper individuals or populations) with a certain appearance (i.e. with certain characters or traits of these) that mate, producing progeny, in which the different traits of the different characters occur in certain relative frequencies.

Now the *basic theory-element of genetics* (T(G)) can be characterized as follows:

$$T(G): = \langle K(G), I(G) \rangle.$$

2.3 The Empirical Claim of Genetics

As we have seen above the claim (I) associated with (1) is not entirely empirical for the theoretical parts of the theory-core inasmuch as *GENOTYPES*, *COMBINATOR* and *DETERMINER* are assumed as given when the claim is made. In the majority of the applications, these components possess a

hypothetical status and the claim depends, therefore, on the corresponding hypotheses. In principle, for any given observational part of the theory core, the number of possibilities for the three components under consideration is infinite. However, as the variety and number of *GENOTYPES* and the mathematical form of *COMBINATOR* and *DETERMINATOR* are determined, in practice, by the application of special hypotheses, or laws, this infinity is restricted. Even such special laws are often not sufficient to determine the theoretical components uniquely. This means that if there is no unique choice of theoretical components as prescribed by the observational part of the theory-core, then an arbitrary set should be taken from the range of possibilities admitted by the theory-core. On the other hand, as we have already said, we use in (I) the expression "ideally fit" in order to indicate that we do not take there into account features of approximation that genetics contains like practically all empirical theories. But, in fact, the empirical claim associated with the theory-element of genetics will always be only approximatively true. Therefore, if we consider approximations and keep in mind the prescription stated above about the theoretical components admitted by the theory-core, we can go from the claim (I) to the following empirical claim:[16]

(II) By adding a set of theoretical components to the non-theoretical ('observational') part of the corresponding theory-core, each given intended system can be approximately extended to, or embedded in, a proper model.

This claim may be trivial if the requirements imposed by the theory-core on the theoretical terms are weak. But this should not be a reason for rejecting the theory-core as trivial. This theory-core serves as a basic theory-core for *all* the intended applications of genetics. Interesting, non-trivial claims, may be obtained by incorporating additional determinations.

[16] For a structuralist approach to features of approximation and a precise formal explication of the notion of the approximative empirical claim, see *Architectonic*, chapter VII.

III. Classical Genetics[17]

Classical genetics is obtained by means of a restriction of the fundamental predicate, consisting in a refinement of it.

Definition 5:
x is a *classical genetics* (x ∈ M(CG)) if and only if

(1) x ∈ M(G)

(2) there exist sets $P_1,..., P_k, F_1,..., F_s$ and functions $DET_1,..., DET_k$ for $i \leq k \leq s$, and $j = 1, 2$ such that

a) each phenotype π of P has the form $\langle p_1,..., p_k\rangle$

b) each genotype γ of G has the form $\langle\langle f_{i1}, f_{i2}\rangle,..., \langle f_{s1}, f_{s2}\rangle\rangle$

c) *DETERMINER* is decomposable.

d) *COMBINATOR* is conservative.

e) the fundamental law of fit holds in one of the two forms (3*) or (3**) for all $r_1,..., r_k$, all $PHENOTYPE\_OF\_PROGENY_1,...,PHENOTYPE\_OF\_PROGENY_k$, all γ, γ', all $\alpha_1,..., \alpha_s, \gamma_1,..., \gamma_s$, all $PARENT_1$, $PARENT_2$, $PROGENY_1,..., PROGENY_k$, and all $PHENOTYPE_1$, $PHENOTYPE_2$ for which *DISTRIBUTOR* is defined, and which occur in x.

(3*) If
- $DISTRIBUTOR(PHENOTYPE_1, PHENOTYPE_2) = \langle r_1 PHENOTYPE\_OF\_PROGENY_1,..., r_k PHENOTYPE\_OF\_PROGENY_k\rangle$
- $DETERMINER(\gamma) = PHENOTYPE_1$
- $DETERMINER(\gamma') = PHENOTYPE_2$
- $COMBINATOR(\gamma,\gamma') = \langle \alpha_1\gamma_1,..., \alpha_s\gamma_s\rangle$

[17] For the following *reconstruction of classical genetics*, we have used the text of (Sinnot, Dunn 1925). This book may be considered the *first textbook of genetics in the Kuhnian sense*, inasmuch as it contains, with pedagogical goals, a clear and actualized - in comparison with the classic (Morgan *et al.* 1915) - exposition of the principles of genetics, paradigmatic applications of them (or 'exemplars'), as well as problems to be solved by the student. We distinguish classical genetics from the so-called 'chromosome theory of (Mendelian) inheritance'. While the first denomination – 'classical genetics' - refers to a theory of heredity which includes no essential links to other theories, the second one – 'chromosome theory of (Mendelian) inheritance' - refers to a theory (an '*inter*field theory', after (Darden, Maull 1977), (Darden 1980), and (Darden 1991)) which includes besides a theory of heredity its interrelations with another body of knowledge, namely: cytology. We think that it is methodologically more adequate to reconstruct first both theories - classical genetics and cell theory - as two distinct and separated theories and just after that to investigate its intertheoretical relations or links (for an analysis of the history-changing 'interfield connections' between genetics and cytology, see (Darden 1991)).

- $j \leq k$

then

$\sum_{\alpha \in C(\gamma, \gamma', j)} \alpha = r_j$

(3**) If
- $PARENT_1$ and $PARENT_2$ are populations
- $MATOR(PARENT_1, PARENT_2) = \langle PROGENY_1,..., PROGENY_k\rangle$
- $j \leq k$
- $APPEARANCE(PROGENY_j) = PHENOTYPE\_OF\_PROGENY_j$
- γ and γ' are *GENOTYPES*
- $DETERMINER(\gamma) = APPEARANCE(PARENT_1)$
- $DETERMINER(\gamma') = APPEARANCE(PARENT_2)$
- $COMBINATOR(\gamma,\gamma') = \langle \alpha_1\gamma_1,..., \alpha_s\gamma_s\rangle$

then

$$\sum_{\alpha \in C(\gamma, \gamma', j)} \alpha = \frac{\| PROGEN_j \|}{\| PROGEN_1 \| + ... + \| PROGEN_k \|}$$

Comments on this predicate:
The central methodology of classical genetics consists in establishing data for the function *MATOR*, i.e. for the probabilities of the different *PHENOTYPES* that occur in the progeny. The genetic hypotheses referring to *COMBINATOR* and *DETERMINATOR* systematize and explain these data.

This restriction of the fundamental model of genetics consists of *an interpretation* of *J*, in *a conceptual extension* of *P* and *G*, in *a so induced refinement* of *APPEARANCE*, *MATOR*, *DISTRIBUTOR*, *DETERMINER*, *COMBINATOR*, as well as of *the specialization* of *DISTRIBUTOR*, *DETERMINER*, *COMBINATOR*, and of the *fundamental law* of genetics.

Classical genetics is concerned with populations, inasmuch as reliable frequencies of traits in the progeny are not obtained through consideration of a single mate. Generally, the progeny of individual parents will not even exhaust all possible phenotypes. The populations are treated as non-empty sets of 'individuals', independently of what might be their nature. The real carriers of phenotypes are individuals, but the manner in which the populations are treated here (without the introduction of the explicit definition in terms of individuals) simplifies the models.[18]

[18] A set of individuals which form the populations - such that each population is a subset of it - may be easily introduced, however. See (Balzer, Dawe 1986).

(2a) The phenotypes are defined as k-tuples of characters. Each phenotype has the form of a tuple $\langle p_1,..., p_k\rangle$, which consists in *component phenotypes*. The usual interpretation of P is the following: each set P_i may be regarded as representing a character and the elements $p_i \in P_i$ as traits or expressions of this character.

(2b) Something similar happens with genotypes. Each genotype is a finite list $\langle g_1,..., g_s\rangle$ of *allelic pairs*, where an allelic pair g_i consists of two *factors*: $g_i = \langle f_{i1}, f_{i2}\rangle$, the factors being the real primitives here. Intuitively, two factors are allelic if they 'work together' in causing or partially causing a particular trait. Formally, this is expressed by putting allelic factors together in one set. We thus obtain a list $F_1,..., F_s$ of sets of factors. For any $i \leq s$ and any two factors f_1 and f_2, f_1 and f_2 belonging to the same F_i means that f_1 and f_2 are allelic. The genotypes in a model thus are defined with respect to a given such list $F_1,..., F_s$ of factor sets. Pairs of factors from the same set F_i we call *genotype components*, so a genotype component has the form $\langle f_{i1}, f_{i2}\rangle$, where $i \leq s$ and f_{i1} and f_{i2} are members of F_i.

Through this characterization of genotypes (in which it is not required that $s = k$, that is, admitting that various pairs of allelic factors may determine one and the same character) it is possible to include in the model both the *interaction of factors* and the so-called *hypothesis of multiple factors*. Note also that the present formalism captures the phenomenon of *multiple allelism*. This phenomenon is given by the fact that different individuals in a species may have different genotype components from the same factor set F_i. We allow for arbitrary, finite factor sets F_i. So, within a species there may be hundred or even thousands of different allelic pairs that are formed from one factor set.

Refinements of the functions (induced by the refinement and the addition of objects):

APPEARANCE assigns k-tuples of characters to the populations.

MATOR is a function that maps pairs of sets of objects (parental populations) in sets of such objects (sets of populations of progeny).

Specializations of laws:

DISTRIBUTOR maps any 'parental' pair of k-tuples of characters into a distribution of phenotypes, which can be written in the following way: $DISTRIBUTOR(PHENOTYPE_1, PHENOTYPE_2) = \langle r_1\pi_1,..., r_k\pi_k\rangle = \langle r_1 PHENOTYPE\_OF\_PROGENY_1,..., r_k PHENOTYPE\_OF\_PROGENY_k\rangle$, where $\sum_{1 \leq i \leq k} r_i = 1$ and each $PHENOTYPE\_OF\_PROGENY_i$ again is a tuple of characters.

As mentioned above, these phenotypes, and their transmission as described by *DISTRIBUTOR*, constitute the data which are theoretically systematized at the level of the genotypes. These data are not purely empirical; the coefficients r_i are not observable: they represent relative frequencies and have to be determined by determining and counting characters and calculating ratios. In CG r_i can be specified in terms of sizes of populations in the offspring: the relative frequency $r_j = \beta_j / \mu$ (that is, phenotypes assigned to $PROGENY_j$ by *APPEARANCE*) is yielded by counting $\beta_j = \| PROGENY_j \|$ (the number of individuals in the j-th population) and taking its ratio to the total number of progeny, $\mu = \| PROGENY_1 \| + ... + \| PROGENY_k \|$.

(2c) *DETERMINER* is a function that maps genotypes into phenotypes such that pairs of allelic factors yields a unique p_i. It is required that *DETERMINER* is decomposable, i.e. that it can be decomposed into a list of 'component functions', one for each component of the phenotypes which occurs. This means that there are functions $DET_1,..., DET_k$, such that each DET_i is a function of pairs of allelic factors into characters: $DET_i(\gamma) = p_i$, and *DETERMINER* is defined as the tuple of all DET_i in the following way:
$DETERMINER(\gamma) = \langle DET_1(\gamma),..., DET_k(\gamma)\rangle$, where the last expression under consideration yields some phenotype $\langle p_1,..., p_k\rangle$.[19]

[19] The condition of decomposability for *DETERMINER* may be defined as follows (see Balzer, Dawe 1997):

DETERMINER is *decomposable* if and only if there exist sets $P_1,..., P_k, F_1,..., F_s$, sets of indices $J_i = \{j(i, 1),..., j(i, \sigma(i))\}$ for $i = 1,..., k$, and functions $DET_1,..., DET_k$ such that:
1) each phenotype π can be represented in the form $\pi = \langle p_1,..., p_k\rangle$ with $p_1 \in P_1,..., p_k \in P_k$

(2d) That *COMBINATOR* is conservative means that the genotypes of any progeny are made up of factors occurring in the genotypes of the parental individuals, i.e. only 'parental' factors may occur in the genotypes of offspring. Recall that *COMBINATOR* operates on two given genotypes of the form $\langle\langle f_{i1}, f_{i2}\rangle,..., \langle f_{s1}, f_{s2}\rangle\rangle$ with factors f_{ij}. If $GENOTYPE_1$ and $GENOTYPE_2$ are represented in the form $\gamma_1^* = \langle\langle a_1, b_1\rangle,..., \langle a_s, b_s\rangle\rangle$, $\gamma_2^* = \langle\langle c_1, d_1\rangle,..., \langle c_s, d_s\rangle\rangle$, respectively, *COMBINATOR* has to produce a distribution $\langle\alpha_1\gamma_1,..., \alpha_s\gamma_s\rangle$, where each γ_i again is a sequence of the form $\langle\langle f_{i1}, f_{i2}\rangle,..., \langle f_{s1}, f_{s2}\rangle\rangle$. If $\langle\langle e_1, f_1\rangle,..., \langle e_s, f_s\rangle\rangle$ denotes an arbitrary γ_i, γ_i occurs in a distribution if and only if:
(i) the factors e_i, f_i are elements of F_i;
(ii) g_i consists only of factors actually occurring in the parental *GENOTYPES* γ_1^* and γ_2^*.

This conservation principle establishes that the genetic material is a stable 'genidentical' entity: in the course of transmission no new factors appear. This expresses something fundamental for the classical model, namely: the purity of the factors. Note that the stronger conservation principle according to which all parental factors must occur in the offspring is not valid. Some parental factors may enter only into combinations that do not get realized in the offspring, and in this sense 'get lost'.

Each genotype being an element of the cartesian product $F = (F_1 \times F_1) \times ... \times (F_s \times F_s)$, and using the notion of a Γ-distribution over F, *COMBINATOR* may be said to be a function in the set $D(F)$ of all Γ-distributions over F: *COMBINATOR*: $F \times F \rightarrow D(F)$. Each value of *COMBINATOR* may be written in the form $\langle\alpha_1\gamma_1,..., \alpha_s\gamma_s\rangle$, where $\alpha_i \in \mathbb{R}$, $\alpha_i \geq 0$, $\sum\alpha_i = 1$, s is some natural number and $\gamma_i \in F$. In this notation the principle of conservation takes the following form: *COMBINATOR* is such that for all γ, γ', $\alpha_1\gamma_1,..., \alpha_s\gamma_s$: if

2) each genotype γ can be represented in the form $\gamma = \langle g_1,..., g_s\rangle$ with $g_1 \in F_1,..., g_s \in F_s$
3) the set $\{1,..., s\}$ of indices is the same as the union of all the sets J_i, $i \leq k$:
$\{1,..., s\} = \cup\{J_i / i \leq k\}$
4) for all $i \leq k$: DET_i maps genotypes into elements of P_i
5) for all $i \leq k$: DET_i properly depends exactly on all its arguments with indices $j(i, 1),..., j(i, \sigma(i))$
6) for all genotypes γ
$DETERMINER(\gamma) = \langle DET_1(\gamma),..., DET_k(\gamma)\rangle$.

That is, each phenotype has the form of a tuple $\langle p_1,..., p_k\rangle$ consisting of *component phenotypes*. Each genotype consists of a tuple $\langle g_1,..., g_s\rangle$ of *component genotypes*. Each sequence $\langle j(i, 1),..., j(i, \sigma(i))\rangle$ picks out the indices of those components of $\langle g_1,..., g_s\rangle$ on which DET_i actually depends, and by 4), DET_i maps the genotype with these components into expressions of the trait P_i.

COMBINATOR $(\gamma,\gamma') = \langle\alpha_1\gamma_1,..., \alpha_s\gamma_s\rangle$ then for all $i \leq s$ all components of γ_i are among the components of γ and γ'. Strictly speaking, the numbers α_i are the values of a Γ-distribution *COMBINATOR*(γ, γ').

(2e) The fundamental law of fit may be formulated in two equivalent ways, one - (3*), where the distribution of genotypes and the relative frequency of individuals in the offspring fit if the sum of all α_i in $C(\gamma,\gamma', j)$ equals r_j - more simple than the other - (3**), which defines the relative frequencies r_j in terms of sizes of populations, the right side of the equation being just the *definiens* for the relative frequency r_j.

3.1. Specializations of Classical Genetics

There are different possible ways of specializing classical genetics. The specializations consist in specifications

a) of the number s of component genotypes,
b) of the concrete mathematical form that *DETERMINER* assumes,
c) of the concrete mathematical form that *COMBINATOR* assumes.

The diverse possibilities of specialization can be partially or totally realized, in an isolated or joint way. In CG, the functions *DETERMINER* and *COMBINATOR* of G have already been specified. However, the specifications introduced were only partial. The process of specification, then, may also be continued in relation to said functions. A specialization in which the three types of specification have been fully realized is denominated *terminal specialization*.

3.1.1 First Line of Specialization:
Equal Probability for all Combinations of Factors ('Mendel's Laws')
Specification of Type c (Mathematical Form of COMBINATOR):

A first line of specialization of CG that characterizes a large class of models concerns *COMBINATOR*; its mathematical form is explicitly postulated: it is assumed that in calculating the genotypes of progeny *all combinations of factors have equal probabilities*. In this way we obtain a specialization which

may be considered as a *general form* of 'Mendel's Laws' insofar as it includes the first as well as the second of 'Mendel's Laws'.[20]

Definition 6:
x is a *classical genetics with equal probabilities* ($x \in M(E)$) if and only if
(1) $x \in M(CG)$
(2) for all *GENOTYPES* $\gamma = \langle\langle a_1, b_1\rangle,..., \langle a_s, b_s\rangle\rangle$, $\gamma' = \langle\langle c_1, d_1\rangle,..., \langle c_s, d_s\rangle\rangle$:
$COMBINATOR(\gamma,\gamma') = \prod_{1 \le j \le s} (¼a_jc_j + ¼a_jd_j + ¼ b_jc_j + ¼b_jd_j)$.

COMBINATOR considers *all possible combinations* of parental factors and assigns each one the *same probability*. To express this, it is necessary to introduce a formal operation of multiplication for distributions of genotypes. The *concatenation* of two tuples $\gamma=\langle x_1,..., x_n\rangle$, $\gamma' = \langle y_1,..., y_m\rangle$, denoted by $\gamma\gamma'$, is defined as the tuple $\langle x_1,..., x_n, y_1,..., y_m\rangle$. It is convenient to abbreviate the distributions of genotypes $\langle\alpha_1\gamma_1,..., \alpha_s\gamma_s\rangle$ as $\sum_{1 \le i \le s} \alpha_i\gamma_i$, or as $\alpha_i\gamma_i +...+ \alpha_s\gamma_s$. The formal multiplication of two distributions of genotypes $\sum_{1 \le i \le s} \alpha_i\gamma_i$ and $\sum_{1 \le i \le s} \beta_i\gamma_i'$, is defined as follows:

$$(\sum_{1 \le i \le s} \alpha_i\gamma_i)(\sum_{1 \le i \le s} \beta_i\gamma_i') = \alpha_1\beta_1\gamma_1\gamma_1' + ... + \alpha_1\beta_i\gamma_1\gamma_i' + ... + \alpha_s\beta_1\gamma_s\gamma_1' + ... + \alpha_s\beta_i\gamma_s\gamma_i'.$$

The iteration of this definition may be obtained by multiplication of the right side of the previous definition with another distribution, etc. The iterated

[20] In the early days of 'Mendelism', what is now labelled 'Independent Assortment' or 'Mendel's Second Law' was not separated from the 'Law of Segregation' or 'Mendel's First Law'. H. de Vries was the first to speak of the 'Law of Segregation of Hybrids' ('loi de disjonction des hybrides' in French and 'Spaltungsgestez der Bastarde' in German) as discovered by Mendel. He spoke of segregation of characters – 'caractères' in French and 'Merkmale' in German - and not of factors or genes, because at that time the distinction between characters or traits on one side and factors or genes on the other side was not clear, see (De Vries 1900). By contrast, another 'rediscoverer', C. Correns, uses the expresion 'Mendel's Law' ('Mendels Regel' in German) to refer to de Vries' 'Law of Segregation' as well as to that which later became 'Mendel's Second Law', see (Correns 1900). The first who used the term 'independent assortment' was T.H. Morgan (Morgan 1913). Only in 1919, Morgan explicitly talked about the two laws, the Law of Segregation and the Law of Independent Assortment of Genes and attributed their discovery to Mendel, referring to them as 'Mendel's First Law' and 'Mendel's Second Law', respectively, see (Morgan 1919). In the standard expositions of classical genetics no general form of the afore-said two 'Mendelian Laws' could be found. They are postulated here on systematic grounds similar to those stated above for the fundamental law of fit.

multiplication of n distributions $\sum_{1 \le i \le s} \alpha_i^j\gamma_i^j$, $j = 1,..., n$ (all of equal "length" s) yields:

$(...((\sum_{1 \le i \le s} \alpha_i^1\gamma_i^1)(\sum_{1 \le i \le s} \alpha_i^2\gamma_i^2))...(\sum_{1 \le i \le s} \alpha_i^n\gamma_i^n))$ or, more concisely:

$\prod_{1 \le j \le n} (\sum_{1 \le i \le s} \alpha_i^j\gamma_i^j)$.

Specification of type a (number s of component genotypes)

By further specializing **E** such that the number s of genotype components is restricted to 1, we obtain the simple case of 'Mendel's First Law' ('the Law of *Segregation*') which is concerned with monohybridism. Obviously, this is a specialization of type *a*.

Definition 7:
x is a *classical genetics with equal probabilities for the four combinations of factors* ($x \in M(O)$) if and only if
(1) $x \in M(E)$
(2) $s = 1$, so that
$COMBINATOR(\langle a_1, b_1\rangle, \langle c_1, d_1\rangle) = (¼a_1c_1 + ¼a_1d_1 + ¼b_1c_1 + ¼b_1d_1)$.

Specification of type b (mathematical form of DETERMINER): terminal specializations:

In the following specializations, in addition to the form of *COMBINATOR* and the number s of component genotypes, the concrete mathematical form of *DETERMINER* is specified.[21] This means that, in this way, *terminal specializations* are reached.

Definition 8:
x is a *classical genetics with equal probabilities for the four combinations of factors* with *complete dominance* ($x \in M(OD)$) if and only if
(1) $x \in M(O)$
(2) for all $i \le k$ there is a set P_i with two elements p_{i1} and p_{i2}: $P_i = \{p_{i1}, p_{i2}\}$

[21] In order to keep things legible the notation $DET_i(f_{i1}, f_{i2}) = p$ is used as an abbreviation for 'for all genotypes $\gamma = \langle\langle f'_{11}, f'_{12}\rangle,..., \langle f'_{s1}, f'_{s2}\rangle\rangle$ such that $\langle f'_{i1}, f'_{i2}\rangle = \langle f_{i1}, f_{i2}\rangle$: $DET(\gamma) = p$'.

(3) for all numbers $i \leq k$ there are exactly two factors f_{i1}, f_{i2}, such that

a) $DET_i(f_{i1}, f_{i1}) = p_{i1}$

b) $\left.\begin{array}{l} DET_i(f_{i1}, f_{i2}) \\ DET_i(f_{i2}, f_{i1}) \\ DET_i(f_{i2}, f_{i2}) \end{array}\right\} = p_{i2}$

Here *complete dominance* (in (3)b) and *recessivity* (in (3)a) are expressed; through them the *phenotypical proportion* 3:1 can be explained. Instances of this specialization correspond to what are presented in the literature as *paradigmatic examples* of 'Mendel's First Law'.[22]

The following specialization specifies *DETERMINER* in another way.

Definition 9:

x is a *classical genetics with equal probabilities for the four combinations of factors* with *incomplete dominance* ($x \in M(OI)$) if and only if

(1) $x \in M(O)$

(2) for all $i \leq k$ there is a set P_i with three elements p_{i1}, p_{i2} and p_{i3}: $P_i = \{p_{i1}, p_{i2}, p_{i3}\}$

(3) for all numbers $i \leq k$ there are exactly two factors f_{i1}, f_{i2}, such that

a) $DET_i(f_{i1}, f_{i1}) = p_{i1}$

b) $\left.\begin{array}{l} DET_i(f_{i1}, f_{i2}) \\ DET_i(f_{i2}, f_{i1}) \end{array}\right\} = p_{i2}$

c) $DET_i(f_{i2}, f_{i2}) = p_{i3}$

In this specialization, the habitual interpretation for P_i and its elements p_{i1}, p_{i2}, p_{i3} is the following: P_i is a character, and its elements are traits or expressions of this character, such that p_{i1} represents a trait, p_{i3} another, and p_{i2} either an *intermediate trait* ('*codominance*') or a *trait different* from the first two ('*overdominance*').[23]

Specification of the type a (number s of component genotypes):

The following specialization concerns, once again - as before O -, the number s of component genotypes, in this case being 2.

[22] See (Sinnot, Dunn 1925), pp. 40-41, 45-50.

[23] See (Sinnot, Dunn 1925), pp. 41-42, 85.

Definition 10:

x is a *classical genetics with equal probabilities for the sixteen combinations of factors* ($x \in M(T)$) if and only if

(1) $x \in M(E)$

(2) $s = 2$, so that

$$COMBINATOR(\langle\langle a_1, b_1\rangle, \langle a_2, b_2\rangle\rangle, \langle\langle c_1, d_1\rangle, \langle c_2, d_2\rangle\rangle) =$$
$$(\tfrac{1}{4}a_1c_1 + \tfrac{1}{4}a_1d_1 + \tfrac{1}{4}b_1c_1 + \tfrac{1}{4}b_1d_1)(\tfrac{1}{4}a_2c_2 + \tfrac{1}{4}a_2d_2 + \tfrac{1}{4}b_2c_2 + \tfrac{1}{4}b_2d_2) =$$
$$\tfrac{1}{16}a_1c_1a_2c_2 + \tfrac{1}{16}a_1c_1a_2d_2 + \ldots + \tfrac{1}{16}b_1d_1b_2c_2 + \tfrac{1}{16}b_1d_1b_2c_2$$

This predicate characterizes what is normally understood by 'Mendel's Second Law' ('the Law of *Independent Assortment*'), inasmuch as it concerns *dihybridism*.[24]

Specification of the type b (mathematical form of DETERMINER): terminal specializations:

In the following specializations it is the concrete mathematical form of *DETERMINER* that is additionally specified.

First, we introduce those specializations in which every pair of factors determines a different character.

Definition 11:

x is a *classical genetics with equal coefficients for the sixteen combinations of factors* with *complete dominance for both pairs* ($x \in M(TD)$) if and only if

(1) $x \in M(T)$

(2) there are two sets P_i, $P_{i'}$ with two elements each: $P_i = \{p_{i1}, p_{i2}\}$ and $P_{i'} = \{p_{i'1}, p_{i'2}\}$ such that

(3) for all numbers $i \leq k$ there are exactly two pairs of allelic factors $f_{i1}, f_{i2}, f_{i'1}, f_{i'2}$, such that

a) $\langle DET_i, DET_{i'}\rangle (f_{i1}, f_{i1}, f_{i'1}, f_{i'1}) = \langle p_{i1}, p_{i'1}\rangle$

b) $\left.\begin{array}{l} \langle DET_i, DET_{i'}\rangle (f_{i1}, f_{i1}, f_{i'1}, f_{i'2}) \\ \langle DET_i, DET_{i'}\rangle (f_{i1}, f_{i1}, f_{i'2}, f_{i'1}) \\ \langle DET_i, DET_{i'}\rangle (f_{i1}, f_{i1}, f_{i'2}, f_{i'2}) \end{array}\right\} = \langle p_{i1}, p_{i'2}\rangle$

c) $\left.\begin{array}{l} \langle DET_i, DET_{i'}\rangle (f_{i1}, f_{i2}, f_{i'1}, f_{i'1}) \\ \langle DET_i, DET_{i'}\rangle (f_{i2}, f_{i1}, f_{i'1}, f_{i'1}) \\ \langle DET_i, DET_{i'}\rangle (f_{i2}, f_{i2}, f_{i'1}, f_{i'1}) \end{array}\right\} = \langle p_{i2}, p_{i'1}\rangle$

[24] See (Sinnot, Dunn 1925), pp. 67-70.

d) $$\left.\begin{array}{l}\langle DET_i, DET_{i'}\rangle\,(f_{i1}, f_{i2}, f_{i'1}, f_{i'2})\\ \langle DET_i, DET_{i'}\rangle\,(f_{i1}, f_{i2}, f_{i'2}, f_{i'1})\\ \langle DET_i, DET_{i'}\rangle\,(f_{i2}, f_{i1}, f_{i'1}, f_{i'2})\\ \langle DET_i, DET_{i'}\rangle\,(f_{i2}, f_{i1}, f_{i'2}, f_{i'1})\\ \langle DET_i, DET_{i'}\rangle\,(f_{i1}, f_{i2}, f_{i'2}, f_{i'2})\\ \langle DET_i, DET_{i'}\rangle\,(f_{i2}, f_{i1}, f_{i'2}, f_{i'2})\\ \langle DET_i, DET_{i'}\rangle\,(f_{i2}, f_{i2}, f_{i'1}, f_{i'2})\\ \langle DET_i, DET_{i'}\rangle\,(f_{i2}, f_{i2}, f_{i'2}, f_{i'1})\\ \langle DET_i, DET_{i'}\rangle\,(f_{i2}, f_{i2}, f_{i'2}, f_{i'2})\end{array}\right\} = \langle p_{i2}, p_{i'2}\rangle$$

Here, *complete dominance for both pairs of factors* is expressed; it is explained through the *phenotypical proportion* 9:3:3:1. Instances of this specialization correspond to what are presented in the literature as *paradigmatic examples* of 'Mendel's Second Law'.[25]

Following the guidelines presented here, we can obtain *other specializations* of E. First there is one in which each genotype component determines a different character (with complete dominance for one pair of factors and partial dominance or codominance for the other). A second one is such that each genotype component determines the same character (with different kinds of dominance and epistasis). This line of specialization characterizes the so-called 'factor interaction'. In other specializations of E the number s of component genotypes or the concrete form assumed by *DETERMINER* is chosen in other ways. For instance, s could be equal to 3 and every genotype component could determine a different character with complete dominance for each genotype component or, with s equal to 3 the genotype components with additive effects would determine the same character varying discontinuously. This latter specialization captures the so-called case of 'multiple factors' for quantitative inheritance. The reader may find the explicit formulation of such specializations elsewhere.[26]

3.1.2. Second Line of Specialization:
No Equal Probability for all Combinations of Factors ('*Linkage*')
Specification of the Type c (*Mathematical Form of COMBINATOR*)

A second line of specialization of **CG** is given by linkage genetics. This concerns *COMBINATOR*; here once again its mathematical form is explicitly specified. The basic idea of linkage genetics - in contrast to **E** - is that not all the

[25] See (Sinnot, Dunn 1925), pp. 63-67.
[26] See (Lorenzano 1995), pp. 143-153.

combinations of factors are equally probable. Some combinations - those of 'linked' factors - are more probable than the others.

In order to express the above, we employ here the definition of linkage genetics given by (Balzer, Dawe 1997), but with the following *modification*. For them, the definition includes the characterization of *COMBINATOR* by means of the recombination frequency for the *loci* i, j as well as gene mapping.

It seems adequate both *historically* and *methodologically* to treat the two separately and to introduce them *in successive steps* - first, linkage genetics, and thereafter, gene mapping. Historically, geneticists had to accept quite a number of cases in which not all the combinations of factors were equally probable,[27] before the idea of a spatial mapping of factors (gene mapping) was postulated and accepted. Methodologically, recombination frequencies can be established without constructing a genetic map but not conversely.

If the parental combinations of factors are so strongly linked that they are always transmitted together, its linkage is denominated *complete*. In such a case, only the original paternal genotypes should be encountered. Usually, however, the linkage is not complete, due to the phenomenon of *recombination*, that is, the phenomenon by which new types of combinations also occur. These new types occur, nonetheless, not as frequently as the paternal types.

In this model, a hypothesis is formulated regarding the linear disposition of the factors along the chromosomes. The linkage expresses a material relation between the factors, whose positions can be changed by means of a physical crossing over occurring during meiosis, between pairs of factors situated in homologous chromosomes.[28]

Linkage occurs in different degrees; it is a quantitative phenomenon, susceptible to measurement in terms of the frequency of crossing over. The *starting point of linkage genetics* is the basic idea that the degree or force of linkage depends on the distance which separates the linked factors ('genes') on a chromosome: the farther two factors are away from each other, the greater is the frequency of crossing over occurring between them.

This can be made precise in the following way:

[27] See (Correns 1900), (Castle 1903), and (Bateson *et al.* 1902, 1905, 1906, 1908). These cases were then regarded as exceptions to what was later called 'Mendel's Second Law'.
[28] But inasmuch as linkage genetics and the recombination frequency, as well as the gene mapping presented below, are based in their totality in genetic analysis (i.e. in breeding data given in our reconstruction by the function *MATOR*) and not in cytological analysis, it is not necessary here to refer to the chromosomes, i.e. to the material objects studied by cytology. The same results are reached without the postulation of any hypothesis about the relationships between genetics and cytology. On the other hand, as we have already said, the treatment of the intertheoretical relations or links of genetics to other theories (cytology, in that specific case) goes beyond the aim of this paper.

Any *GENOTYPE* $\gamma = \langle\langle a_1, b_1\rangle,..., \langle a_s, b_s\rangle\rangle$ has a natural order built in, given by the indices 1,..., s. Furthermore, any *GENOTYPE* is composed of two strands, defined as follows

> If $\gamma = \langle\langle a_1, b_1\rangle,..., \langle a_s, b_s\rangle\rangle$ is a *GENOTYPE*, the two *strands of* γ are given by the tuples $\langle a_1,..., a_s\rangle$ and $\langle b_1,..., b_s\rangle$.

Since the factors in $\gamma = \langle\langle a_1, b_1\rangle,..., \langle a_s, b_s\rangle\rangle$ are ordered linearly by their indices, such indices can be simply conceived as their *positions* or *loci*. Thus, every index $i \leq s$ stands for one *position*, often 'occupied' by different pairs $\langle a_i, b_i\rangle$, $\langle c_i, d_i\rangle$ for which a_i, b_i, c_i, d_i belong to the factor set F_i.

The set of *LOCI* of a model x is defined as the set of these positions:

$$LOCI(x) = \{1,..., s\}.$$

The *central hypothesis of linkage genetics* now takes the following form: the farther away two *loci* are on a genotype, the greater is the frequency of crossing over occurring between these *loci*, where distance of loci is simply given by their numerical difference.

In order to formalize the notion of frequency of crossing over (or *recombination frequency*) we first introduce the concept of a *new strand.*

If γ, γ' are *GENOTYPES* and i, j *loci*, then the following is defined:

> A strand s is *new with respect to* γ, γ', i and j if and only if
>
> 1) s has the form $\langle e_1, e_s\rangle$
> 2) γ and γ' have the form $\langle\langle a_1, b_1\rangle,..., \langle a_s, b_s\rangle\rangle$ and $\langle\langle c_1, d_1\rangle,..., \langle c_s, d_s\rangle\rangle$, respectively
> 3) $\langle e_1, e_s\rangle$ is different from each of the four pairs: $\langle a_i, a_j\rangle$, $\langle b_i, b_j\rangle$, $\langle c_i, c_j\rangle$, $\langle d_i, d_j\rangle$.

The *GENOTYPE* γ^* is *new with respect to* given genotypes γ, γ' and *loci* i, j, if at least one of the two strands of γ^* is new with respect to γ, γ', and i, j.

The frequency of a crossing over between two *loci* can be defined in two stages. The probability coefficient which in the genetic distribution $COMB(\gamma, \gamma')$ is associated with genotype γ^* gives us the 'frequency of occurrence' of γ^* in CG. The frequency of γ^* is just α_i if $COMB(\gamma, \gamma')$ has the form $\sum \alpha_i\gamma_i$ and γ^* is γ_i. A new genotype will be 'observed' if crossing over has taken place, i.e., by means of extra assumptions about *DETERMINER* this inference can be made from observations of phenotypes. If γ^* is just this new genotype, its frequency

can be obtained in CG in the way described above. This frequency informs how often crossing over resulting in γ^* has occurred. The coefficient of γ^* in the genotype distribution of progeny provides the frequency of crossing over between *loci* i and j for given parental genotypes γ, γ', resulting in genotype γ^*. We define:

> If γ^* is new with respect to γ, γ', i and j, then the *recombination frequency in loci* i, j *of* γ and γ' *resulting in* γ^*, $RCF(\gamma,\gamma',\gamma^*,i,j)$, is defined by $RCF(\gamma, \gamma', \gamma^*, i, j) = \alpha_k$, where α_k is the coefficient occurring with γ^* in $COMB(\gamma, \gamma')$.[29]

In this definition, *RCF* is defined relative to a given genotype γ^* in the offspring. We obtain the 'absolute' recombination frequency for the parental genotypes γ, γ' and *loci* i, j by considering all the possibilities of recombination, calculating the corresponding frequencies, and adding them up.

In this way the *recombination frequency* for γ, γ' *in loci* i, j is obtained:

$$RCF(\gamma, \gamma', i, j) = \sum RCF(\gamma, \gamma', \gamma^*, i, j),$$

where summation is over all the genotypes γ^* which are new with respect to γ, γ', i and j. Now, this new line of specialization of CG can be characterized:

Definition 12:

x is a *linkage genetics* ($x \in M(L)$) if and only if

(1) $x \in M(CG)$

(2) for γ, γ' in *loci* i, j: $RCF(\gamma, \gamma', i, j) = \sum_{\gamma^*,\, new} RCF(\gamma, \gamma', \gamma^*, i, j)$.

3.1.2.1. Refinement of Linkage Genetics: Gene Mapping

Linkage- or *genetic maps* are constructed through the use of *recombination frequencies*. These maps are representations of *loci* on the real number line such that the order and distances of the representing numbers homomorphically represent the order and distances appearing in the genetic material. By comparing the measured distances and fitting them, so as to make sense of additivity, the order is established.

A *genetic map* for a model x of CG is defined as a function

h: $LOCI(x) \rightarrow \mathbb{R}$,

subject to the following requirements:

[29] Recall that $COMB(\gamma, \gamma')$ has the form $\langle\alpha_1\gamma_1,..., \alpha_S\gamma_S\rangle$ and γ^* is one of $\gamma_1,..., \gamma_S$.

1) for all $i \in LOCI(x)$: $h(i) \geq 0$
2) for all *GENOTYPES* $\gamma, \gamma' \in x$ and all $i, j \leq s$
$100 \cdot RCF(\gamma, \gamma', i, j) = |h(i) - h(j)|$.[30]

The determination of a genetic map proceeds by a kind of process of trial-and-error. Initially, an assumption is stated concerning the order. This is followed by fixing some h-values through experiments of recombination. If these values fit with the assumption about the order, it is accepted. But if they don't fit, the assumption about their order must be modified and begun over again. The usual procedure in determination of h-values involves the determination of as many *recombination frequencies* as possible. The h-values may be determined from the equations in 2) and, if they are correct, the order falls out naturally.

Having characterized the function h, the same can be done with *gene mapping*, which consists in a *conceptual extension* of linkage genetics (the objects on both sides are identical; only the new function h is added), and in a *specialization* of its laws (given by the axioms for the genetic map).

Definition 13:
x' is a *gene mapping* ($x' \in M(M)$) if and only if there is x and h, such that
(1) $x' = \langle x, h \rangle$
(2) $x \in M(L)$
(3) h: $LOCI(x) \rightarrow \mathbb{R}$ is a genetic map for x.[31]

IV. The Structure of Genetics: Its Theory-Net

The structure of genetics may be represented as a theory-net, where the nodes are given by the different theory-elements, and the edges represent different relations of refinement and specialization (Fig. 2). A refinement of G, which has not been addressed here, characterizes molecular genetics. Another, which constitutes the subject of this article, characterizes classical genetics CG, whose principal lines of specialization are given by E and L.

In the latter the *only node* involved *in a refinement relation* is M. Here, a new function - the genetic map h - is added to those of CG. The *other* nodes,

[30] The factor 100 is inserted in 2), in order to obtain percentages rather than relative frequencies.
[31] For a formulation of a special constraint for gene mapping, according to which different models representing cases from the same species are required to have identical genetic maps, see (Balzer, Dawe 1997), pp. 84-87.

except CG of course, are related to the nodes appearing immediately on top of them by a *specialization* relation.

These are not, of course, the only possible restrictions of the set-theoretic predicate that characterizes M(G).[32] They are sufficient, however, for characterizing the *principal lines of specialization and refinement of classical genetics*.

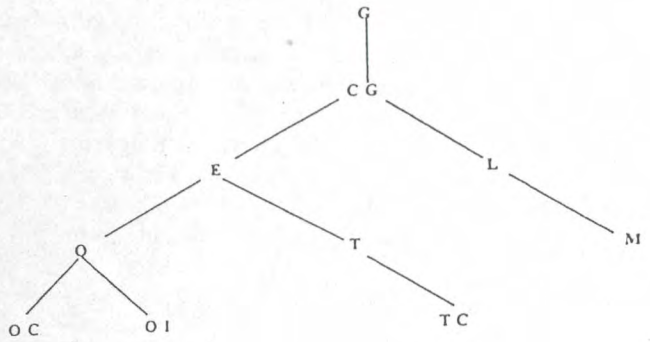

Fig. 2

V. Conclusions

In the present article, a *reconstruction of genetics* has been proposed, focusing on *classical* genetics. This reconstruction has been realized with the instruments, duly modified and extended in accordance with the considered case, of the *structuralist conception of theories*.

In this reconstruction, *genetics exemplifies all the characteristics* that have been considered as *essentials of empirical theories* in general according to this conception. Thus, it can be stated that this reconstruction constitutes a *successful application* of that metatheory. In particular, we may claim to have isolated the fundamental law of genetics, namely: the law of fit. This claim

[32] We could, for example, reconstruct mutations by means of a modification of *COMBINATOR* such that the principle of conservation is given up. Such an analysis, however, goes beyond the limits of the present article.

contrasts with Smart's and Kitcher's claims, respectively, that in biology in general and genetics in particular no fundamental law can be made out.

This article presents a *more precise and complete reconstruction of classical genetics* than has been presented to date; nonetheless, we recognize the incompleteness of the current reconstruction, at least in the following three respects: first, the reconstruction offered here has been *limited to the synchronic aspects* of genetics and does not encompass its diachronic aspects. Secondly, this reconstruction has been *limited to the most significant refinements and specializations of classical genetics*. And, finally, this article *does not incorporate the analysis of intertheoretical relations* in genetics, in particular, the important history-changing relations with cytology, on one hand, and, insofar as molecular genetics may be considered, with biochemistry on the other hand.

Nevertheless, we hope that we have contributed to paving the way for further developments in these directions.

Universidad Nacional de Quilmes
e-mail: pablol@unq.edu.ar

REFERENCES

Asquith, P.D., Giere, R.N. (eds.) (1980). *PSA 1980*. East Lansing, Michigan: Philosophy of Science Association.

Balzer, W., Dawe, C.M. (1986). Structure and Comparison of Genetic Theories I & II. *British Journal for the Philosophy of Science* 37, 55-69, 177-191.

Balzer, W., Dawe, C.M. (1997). *Models for Genetics*. Frankfurt am Main: Peter Lang.

Balzer, W., Lorenzano, P. (2000). The Logical Structure of Classical Genetics. *Journal for General Philosophy of Science*, 1-24.

Balzer, W., Moulines, C.U., Sneed, J. (1987). *An Architectonic for Science*. Dordrecht: Reidel.

Bateson, W., Saunders, E.R. (1902). Experimental Studies in the Physiology of Heredity. *Reports to the Evolution Committee of the Royal Society*, Report I.

Bateson, W., Saunders, E.R., Punnett, R.C., Kilby, H. (1905). Experimental Studies in the Physiology of Heredity. *Reports to the Evolution Committee of the Royal Society*, Report II.

Bateson, W., Saunders, E.R., Punnett, R.C. (1906). Experimental Studies in the Physiology of Heredity. *Reports to the Evolution Committee of the Royal Society*, Report III.

Bateson, W., Saunders, E.R., Punnett, R.C. (1908). Experimental Studies in the Physiology of Heredity. *Reports to the Evolution Committee of the Royal Society*, Report IV.

Castle, W.E. (1903). Mendel's Law of Heredity. *Science* 18. 396-406.

Correns, C. (1900). Gregor Mendels Regel über das Verhalten der Nachkommenschaft der Bastarde. *Berichte der Deutschen Botanischen Gesellschaft* 18. 158-168.

Darden, L. (1980). Theory Construction in Genetics. In: Nickles, T. (ed.), (1980), 151-170.

Darden, L. (1991). *Theory Change in Science. Strategies from Mendelian Genetics*. Oxford: Oxford University Press.

Darden, L., Maull, N. (1977). Interfield Theories. *Philosophy of Science* 44, 43-64.

Dawe, C.M. (1982). *The Structure of Genetics*. Doctoral Dissertation. London: University of London.

Dawe, M.S., Dawe, C.M. (1994). *Prolog for Computer Science*. London: Springer.

Goodenough, U., Levine, R.P. (1974). *Genetics*. London-New York-Sidney-Toronto: Holt Rinehart & Winston.

Henkin, L., Suppes, P., Tarski, A. (eds.) (1959). *The Axiomatic Method*. Amsterdam: North-Holland.

Kitcher, P. (1984). 1953 and All That: A Tale of Two Sciences. *The Philosophical Review* 93. 335-373.

Lindenmayer, A., Simon, N. (1980). The Formal Structure of Genetics and the Reduction Problem. In: Asquith, P.D., Giere, R.N. (eds.), (1980), 160-170.

Lorenzano, P. (1995). *Geschichte und Struktur der klassischen Genetik*. Frankfurt am Main: Peter Lang.

Morgan, T.H. (1909). What are Factors in Mendelian Inheritance?. *American Breeders' Association Report* 6. 365-368.

Morgan, T.H. (1913). *Heredity and Sex*. New York: Columbia University Press.

Morgan, T.H. (1919). *The Physical Basis of Heredity*. Philadelphia: Lippincott.

Morgan, T.H., Sturtevant, A.H., Muller, H.J., Bridges, C.B. (1915). *The Mechanism of Mendelian Heredity*. New York: Henry Holt and Company.

Nickles, T. (ed.) (1980). *Scientific Discovery: Case Studies*. Dordrecht: Reidel.

Rizzotti, M., Zanardo, A. (1986). Axiomatization of Genetics I & II. *Journal of Theoretical Biology* 118, 61-71, 145-152.

Sinnot, E.W., Dunn, L.C. (1925). *Principles of Genetics: An Elementary Text, with Problems*. New York: McGraw-Hill; 2nd ed., 1932; 3rd ed., 1939; with T. Dobzhansky as co-author, 4th ed., 1950; and 5th ed., 1958.

Smart, J.J.C. (1959). Can Biology be an Exact Science?. *Synthese* 2, 1-12.

Smart, J.J.C. (1963). *Philosophy and Scientific Realism*. London: Routledge and Kegan Paul.

Sneed, J.D. (1983). Structuralism and Scientific Realism. *Erkenntnis* 19, 345-370.

Strickberger, M.W. (1985). *Genetics*. New York-London: Macmillan, 3rd ed.

Vries de, H. (1900). Das Spaltungsgesetz der Bastarde. *Berichte der Deutschen Botanischen Gesellschaft* 18. 83-90.

Woodger, J.H. (1959). Studies in the Foundations of Genetics. In: Henkin, L., Suppes, P.,Tarski, A. (eds.) (1959), 408-428.