EGON BRUNSWIK

PERCEPTION AND THE REPRESENTATIVE DESIGN OF PSYCHOLOGICAL EXPERIMENTS

UNIVERSITY OF CALIFORNIA PRESS
BERKELEY AND LOS ANGELES : 1956

University of California Press
Berkeley and Los Angeles, California
Cambridge University Press
London, England

Second edition, revised and enlarged

Library of Congress Catalog Card Number: 56-6984

Printed in the United States of America
By the University of California Printing Department

PROFESSOR EGON BRUNSWIK died unexpectedly a short time after finishing the manuscript of this book. It is fortunate indeed for psychology that this work was completed before the author's death. With the addition of Part Two, the book represents a comprehensive and integrated exposition of Brunswik's theoretical approach. He was a true pioneer who was constantly striving to break out of the confines of orthodoxy and to develop a new methodology which would increase the scope and validity of psychological research. This book is an impressive demonstration of the success of his creative efforts.

The friends and colleagues of Professor Brunswik at Berkeley were anxious for the publication of this book to go forward without delay. They undertook the task of seeing the book through the final stages of publication. As the book goes to press, they are more convinced than ever that Brunswik's creative genius will have an ever increasing impact on the development of psychology.

PREFACE

THIS BOOK has been written with two major purposes in mind. One is the exposition of the more complex attainments of perception, those attainments that help stabilize our grasp of the relevant features of the physical and social environment. The other purpose is the development of the only methodology by which the desired content can be reached, that is, representative design. Together they lay the foundation for a functional approach to perception that stresses biological adjustiveness; this approach supersedes both traditional psychophysics with its oversimplified peripheralism and Gestalt psychology with its undue encapsulation into the internal dynamisms of the cognitive process.

In our presentation the sequence of content and method is reversed. Part One develops the argument for a shift from systematic to representative design as the necessary methodological adjunct to the functional approach with the use of no more than a linear string of examples; these examples run from classical psychophysics through Gestalt illusions to the perceptual constancies and their counterpart in the perception of social objects and also include some applications in the field of learning. This first part has appeared previously, under its present title, both as a contribution to the *Berkeley Symposium on Mathematical Statistics and Probability* (ed. by Jerzy Neyman, 1949) and, with different paginations, as a preprint (1947; reissued, with minor corrections, 1949); it still must be considered the core of the presentation. We have therefore considered it appropriate to call this book a second edition. The only change in Part One is in the year of a reference to one of the writer's publications (1952) which had been repeatedly postponed (and also appeared under a different title from the one listed in the earlier edition).

Part Two branches out from those two of the four model experiments in Part One in which the level of complexity has reached the criterion of functionality, expanding their basic principles over other areas of perception. The writer has attempted to develop the functional study of perception from the beginnings of his academic career, first at Vienna and later at California, and there is strong but by no means exclusive emphasis on his work and that of his collaborators. A good proportion of the forty added illustrations has previously appeared in German or American periodicals or in the writer's *Wahrnehmung und Gegenstandswelt* (1934); some of the others are new or are based on unpublished dissertations. All but a few of the materials are for the first time made conveniently accessible to the American and English reader.

The first three of the eight new chapters (§§XI to XIII) expand on the central topic of the perceptual constancies while the fourth (§XIV) investigates their relation to thinking. Social perception is again taken up in the fifth (§XV). In selecting the materials for these five chapters advantage has been taken of what had become increasingly clear to this writer within the last several years, that is, that hybrid designs combining features of both systematic and representative design are likely to continue and even to increase in frequency within the near future. The first five chapters of Part Two all move in such an intermediate area, either by the use of an informal type of situational sampling which we have called "canvassing," or by superimposing representative features upon systematic factorial design. At the same time essential limitations in the mathematical tool geared to factorial design—analysis of variance—will be pointed up.

Representative design, being a more radical departure from traditional practices of experimentation than is factorial design, makes for difficulties in execution. Representative design in its full scope requires not only a basic theoretical and methodological restructuring but is a formidable task in practice as well. Ideally, it would take concerted research projects of a magnitude hitherto unheard of in experimental psychology. These would tackle a host of problems simultaneously and require the development of new mathematical methods for evaluation. The few inroads reported in the first edition are essentially in the character of methodological demonstrations and barely scratch the surface of the problem. An example of a fully representative design, and one which lies exclusively within textural ecology, will be presented in the sixth of the new chapters (§XVI). This design expands the fully representative designs of Part One (social perception, §VI; and size-constancy, §VII/3) in the direction of psychological research without a subject, a case which is as important theoretically as it is didactically.

Probabilistic cue learning, which is the mediational corollary of distal attainment as dependent on changing ecologies, is taken up next (§XVII). The subsequent chapter (§XVIII) is an excursion into clinical applications; it was prompted by the unexpectedly lively response to the first edition that has come from these quarters. The book concludes with some theoretical considerations.

Help in statistical matters has been generously given by Drs. Rheem F. Jarrett and Robert Rollin. Dr. Leo Postman has been ready with advice on matters too varied and occasions too numerous to count. Mr. Stanford E. Seidner has kindly consented to the use of some of the results of his doctoral dissertation which is still in progress.

EGON BRUNSWIK

Berkeley, California
February, 1955

CONTENTS

PART ONE

SYSTEMATIC AND REPRESENTATIVE DESIGN OF PSYCHOLOGICAL EXPERIMENTS:

WITH RESULTS IN PHYSICAL AND SOCIAL PERCEPTION

PART TWO

PERCEPTION: THE ECOLOGICAL GENERALITY OF ITS DISTAL AIM

PART ONE

SYSTEMATIC AND REPRESENTATIVE DESIGN OF PSYCHOLOGICAL EXPERIMENTS

WITH RESULTS IN PHYSICAL AND SOCIAL PERCEPTION

I

Introduction. Differential and Functional Problems in Psychology

Science has a way of growing in spearheads. Ever since Galton and Pearson established correlation statistics at the end of the last century, in close connection with problems of heredity and "individual differences" of anthropometric and psychological traits, the special field of differential psychology has supplied the content in terms of which psychologists could develop, or absorb, a general methodology of statistical evaluation.

Meanwhile, experimental psychology, with freedom of design added to freedom of evaluation, complacently took for granted what will be characterized, later in the present discussion, as "classical" design. Nineteenth-century experimental "methods," concentrating on psychophysics (Fechner, 1860), memory (Ebbinghaus, 1885), and related functional or "stimulus-response" problems, did not challenge the principles underlying this approach; they are but specific elaborations of procedure and evaluation within an accepted framework. General experimental methodology in psychology thus remained, at least as far as explicitly verbalized, programmatic statements are concerned, virgin territory up to the time of the first impact of Fisher's factorial design upon psychology less than a decade ago.

But whereas factorial design departs from classical notions by becoming multivariate, it does not in itself guarantee a second important feature, to be called "representativeness" of design. The main purpose of this paper is to demonstrate the feasibility—and in fact the informal existence in a limited way for some time—as well as the requiredness of this second change of policy in addition to mere multidimensionality, at least if the further development of experimental psychology is to be toward a more truly "functional," "molar" rather than "molecular," dynamic rather than static science.[1]

Although it is claimed that the arguments involved are valid for the entire domain of experimental psychology, the special field of perception is singled

[1] For a survey of psychological "systems" see Heidbreder (1933). More recent summaries and selected bibliographies are to be found in Brunswik (1946*b*, 1952). For the sake of brevity, all stimulus-response problems will here be designated as functional problems—in contradistinction to the differential problems mentioned above—regardless of whether or not their treatment is on the adequate level of complexity characterizing the "functional approach" as at least vaguely anticipated by what is known as the school of Functionalism (see also fig. 7, including the legend).

out for purposes of illustration. Not only is the psychology of perception older than that of such other fields as learning or motivation, but it also has served as a pacemaker in the way of decisive changes in basic outlook. Perception thus may be best suited as a paradigm not only for the past but also for things likely to come. Some of them are already discernible in other fields possibly more crucial with respect to content but less advanced methodologically.

Within the field of perception, the special topic of the perception of linear magnitude seemed to provide the best available series of examples to the point. A model-sequence of four experiments (see also Brunswik, 1946*a*) has been chosen to represent four distinct stages of development. Experiments A to C involve the perception of physical magnitude under conditions of increasing complexity, whereas experiment D further includes personality variables in a study of social perception.

This discussion will be followed by a more generalized survey of the possible uses of basic measures of concomitant variation such as the correlation coefficient. Special emphasis will thereby be given to the concepts of reliability and validity as they may refer, over and above their more customary differential psychological application, to functional stimulus-response patterns, the traditional domain of the systematic experiment. It seems that, in the present spadework stage at least, one may in this process get by with no more than the most elementary statistical concepts, which is all the present writer is familiar with. The analysis of the chosen historical sequence of experiments with increasing representativeness of design will be preceded by a brief listing, pointed for use with our examples, of the types of variables within the scope of the psychologist and of some fundamental alternatives in the designing of psychological research.

I. The Variables Entering Objective Psychological Research

In any kind of research endeavoring to find general regularities, if not "laws," discussion is adequate only if carried out in terms of well-defined broader categories defining "variables" or "dimensions" rather than in terms of individual occurrences. Examples of such variables are "length," "area," "intelligence quotient (IQ)," etc. While the basic character of variables must be considered the same in all objective sciences, including modern psychology, in that they are defined, directly or indirectly, by "operations" of the "physicalistic" type such as, primarily, measurement,[2] specifications are nonetheless in order within the framework of the various disciplines. The subsequently listed classifications of variables seem those most urgent from the standpoint of the perception psychologist. A survey of some of them is given below in the comprehensive scheme presented in figure 7.

1. Classification with Respect to Regions Relative to an Organism

Depending on whether a series of events is studied with reference to the time before, while, or after it plays upon an organism, one may distinguish

[2] For a discussion of the objective vs. the subjective approach in psychology, of "behaviorism" vs. "introspectionism"—perhaps the one issue of general methodology in psychology that supersedes that of research design—see the sources mentioned in the footnote to the Introduction, especially Brunswik (1952). See also below, § VI/3.

external *stimulus* variables (in a relatively broad use of the word), *S*, *organismic* variables, *O*, and behavioral *response* variables, *R*. Depending on the inclinations of the researcher, overt motor responses as well as their further "results" or "effects," including written protocol marks and similar indications of perceptual judgments or "estimates," may or may not be taken to reflect subjective, "conscious" impressions.

Many of the environmental stimulus variables mentioned by psychologists, such as "physical size" or "physical color," seem at first glance simply to be taken over from physics or chemistry. Others, such as "food," "sit-uponableness" (William James), "likability of a person," etc., are obviously conceived with an eye to potential effects upon organisms. In both cases the "dispositional" character of the definition (Carnap) is maintained, the psychological slant of the latter type of variables notwithstanding. Upon closer inspection, however, even the former often reveal psychological entanglement when they appear in the context of a psychological experiment. For example, the "sizes" of physical objects (more precisely, of physicist's objects) figuring as one of the major stimulus variables in the statistical survey of size constancy (see § VII/3) are in fact to be specified as "sizes of objects of attention, i.e., of potential manipulation or locomotion, of a certain human being."

Considering the arbitrariness with which objects, spaces between objects, or parts thereof may be subjected to physical measurement, sizes *per se* can hardly be thought of as possessing a finite range or distribution. Sizes attended to in perception or behavior, however, although likewise strictly determined by external measurement (and not to be confused with perceptual size estimates) delimit a much more closely circumscribed reference class or "universe" of sizes; we must therefore not be surprised if they show a definite central tendency. Actually, their logarithms even seem to tend toward a normal distribution (see Brunswik, 1944, fig. 1). This feature is of the utmost importance in the present context, since it allows, and in fact demands, an application of the principles of representative sampling to variables which at face value may not appear capable of being sampled. It is the type of organism-centered specifying redefinition mentioned above which may be summarized by saying that *stimulus variables are "ecological"*[3] rather than purely "physical" or "geographic" in character.

The terms employed in delimiting ecological universes are sometimes quite similar to those well known from the sampling theory of individual differences. An example is David Katz's (1911/1935) introduction of the concept of "normal" conditions of illumination, defining a considerable range within which certain statements about the perception of object colors are to be understood to hold.

Stimulus variables may be subdivided "regionally"—as likewise may behavioral responses and their effects in contexts other than the present—into

[3] Concerning the use of the term "ecological" in psychology, borrowed from botany and zoölogy to designate the natural or customary habitat, or surrounding universe, of a species, culture, or individual, with all its inherent variation and co-variation of factors, see Brunswik (1943). Use of this term in psychology was first suggested by Lewin (1943), although in a technically less specified, broader sense roughly encompassing what is more commonly known as stimulus-response problems (in contradistinction to intra-organismic problems of "field" dynamics). See also § VI.

distal, S_D, and *proximal*, S_P (see Heider, 1939; Brunswik, 1939*a*). Distal variables are defined independently of effects or representation on the sensory surface of an organism in a certain situation; proximal variables—in the present paper treated as if they would vary in unison with the primary responses within the organism, the "peripheral sensory excitations"—are defined as effects (originating in distal conditions) as they impinge upon the sensory surface of a subject. An example of a proximal stimulus variable is given by the (absolute or relative) sizes of retinal projections, whereas the sizes of the manipulable physical bodies in the environment which together with distance determine these projections exemplify a distal stimulus variable.

Environmental conditions even more remote may be called *covert distal* or second-order distal variables, S_C; they are exemplified by the IQ's, "objective" likabilities, etc., of "social objects" viewed by an organism, whereas such physiognomic features as the heights of the foreheads of these social objects may be taken as an ecologically further specified subclass of the above-mentioned overt distal variable, bodily size.

Of the organismic variables, we will single out the *central* (as contrasted to the [sensory and motor] peripheral) variables for special consideration. They may be subdivided into relatively permanent traits varying from individual to individual in a population, and relatively transient sets, attitudes, or motivational states and other dispositions varying within the individual from one time to another. They shall be designated by lower-case vowels, the former by u (for the middle vowel in "population"), the latter by a (for the initial vowel in "attitude").

2. CLASSIFICATION WITH RESPECT TO ROLE WITHIN A PATTERN OF VARIABLES

Experimental variables constituting the initial conditions of research are customarily called *independent* variable(s), x. Final observed effects are the *dependent* variable(s), y.

Links of the causal chains connecting x and y will be said to define the *mediating* variables, M. An example is retinal size as an intermediary between physical object size, on the one hand, and perceptual estimates, on the other. The accessories to x in a field, e.g., the area adjacent to lines which are to be compared with respect to length (see fig. 2) shall, for want of a better term, be called neighboring variables, N.

The distinction between M and N is not easy and in the end hinges upon a clarification of the relationship of correlation to causation; a casual distinction in terms of the examples mentioned may, however, suffice for the purposes of the present paper.

3. CLASSIFICATION WITH RESPECT TO FUNCTIONAL VALIDITY OF A RESPONSE

This is a classification with respect to the mechanisms inherent in organismic "functioning" that lead to *dependency in a statistical sense*. The term validity may be readapted from the statistics of individual differences to indicate how much one event justifies expectation, under representative conditions, of the occurrence of another event pertaining to another variable or region within or about a given individual (see § VI/3). Considering the specific dynamic interaction (see § IV/1), on the one hand, and the stabilization mechanisms (see § V/1) characteristic of patterns of life, on the other, both neighboring and mediating conditions may be *contributing*, or, at least approximately, *noncontributing* to the response within a given type of functional context.

II. Fundamental Aspects of Design in Psychological Research

The preparation for the gathering of factual evidence may be viewed under three aspects: choice of variables, manner of their variation, manner of their concomitant variation.

1. CHOICE OF VARIABLES ENTERING THE SCOPE OF AN EXPERIMENT

One of the most fundamental issues in the choice of variables is the *number of variables* allowed to enter the scope of an experiment. Whereas the classical ideal is that of the "rule of one variable" (see Woodworth, 1938, pp. 2 f.), modern trends in psychology are toward multiple variable design, which makes successful handling of problems of "context" possible. Although the presence of many variables in a situation has always been recognized, the number of those considered, i.e., specifically controlled or evaluated in computation, has in the past usually been drastically limited; this was often done under the assumption that the remaining variables were irrelevant (see § III/1).

Of equal if not greater importance is *regional reference* in the sense introduced above (§ I/1). As will be seen, research in perception was first predominantly proximal-peripheral but is now becoming increasingly central-distal in its approach. This further implies an increasing *reach* of the "arch" spanning between the variables considered (for both trends see below, fig. 7). If mediation is to be included within its scope, as in certain ways it should, distal emphasis, then, implies multivariate design. However, in the study of individual differences, especially in that of the inheritance of traits from one generation to another, mediation is usually left unscrutinized within wide limits, with the result that the entire pattern of approach may remind one of the "one independent variable—one dependent variable" scheme of classical stimulus-response research, in spite of the enormous number of unknown conditions filling the gap between the two variables observed.

2. MANNER OF VARIATION OF THE VARIABLES CHOSEN

This refers to the frequency distribution of the values of a variable along a scale. The major alternatives are listed subsequently.

Graphic symbols for each of them as well as for the other methodological concepts discussed in this paper are introduced in figure 7. Some of the "modifiers" apply especially to perception; but they may easily be augmented, by analogy, for overt behavior.

a) *Formalistic, "systematic," policies, traditionally associated with active control by the experimenter.*—In this type of procedure, variation is dictated by an experimenter in a laboratory situation and patterned after arbitrary, usually formalistic, i.e., clean-cut, principles possessing a certain symmetry and regularity. It covers both those cases in which variation is allowed to occur and those in which it is not permitted.

The former case shall be called *systematic variation* (in the narrower sense of the word). In the typical case the values of the variable in question are spaced in even, discrete steps of equal frequencies, and along an arbitrary range of the variable in question resulting, say, in a rectangular distribution of values.

The opposite alternative to systematic variation within the general framework of systematic design is the *holding constant* of a condition. This can be done either by maintaining a distinct finite value other than zero, or by minimizing or nullifying, i.e., reducing to near zero or to zero. These latter alternatives can often not be clearly distinguished, e.g., when an even black background is used to eliminate the possibility of disturbance issuing from the surroundings.

The chief advantage to be gained from this technique, envisaged in J. S. Mill's "method of difference," is the exclusion of a condition as a possible contributor to variations in the response which then, if present, must be attributed to other causes. It is in this sense that variables held constant do enter the scope of the experiment in question. But they do so only in a negative way, without actually being given a chance as a potential factor.

The frequent specific case that *two or more discrete constant values* with no clear-cut specification of their interval (e.g., two different instructions to the subject, or two or more different ink blots in the well-known Rorschach test) are presented is an intermediary case between systematic variation and the holding constant of a condition.

b) *Representative variation and passive control.*—In this second group of alternatives, the values of the variable in question are left alone and merely registered in their entirety in the passive procedure of observational control of the actuarial type. Or else they are interfered with merely to the point of extracting a sample that is more or less "representative" of the entire reference class in question. This may be left to chance in random sampling; or a more active attempt to assure representativeness may be made by what is known as controlled sampling, a procedure which, as will be seen, is not free from the danger of arbitrariness when applied to the sampling of stimulus situations (or tests) in canvassing an experimental problem.

3. MANNER OF CO-VARIATION OF THE VARIABLES AMONG ONE ANOTHER

For the handling of concomitant variation, or, in short, of "co-variation" among variables in an experiment there are again the two major possibilities of systematic and of representative design.

a) *Systematic co-variation.*—The following three subvarieties of this first alternative shall be distinguished here:

Artificial tying of variables. Suppose, as is the case in the Galton-bar experiment (Experiment A, fig. 1), that two lines at the same distance from the eye are to be compared with each other. Owing to the equality of distance, projections on the retina of the eye (variable *P*) are equal when "bodily" lengths of the two half-bars (variable *B*) are equal; in short, the two "points of objective equality" (POE's) coincide. The two stimulus variables are thus inseparably tied by arrangement of the experimenter. This holds true as long as the scope of the investigation is confined to the laboratory situation in question. The two variables then vary in perfect unison; their correlation is artificially made to be 1. Whether the response is specific to the one or the other variable, or is a function of both, thus becomes indistinguishable.

For the particular variables chosen as examples the state of affairs may be specified as "channeled mediation."

In experimental procedure, artificial tying of variables makes it possible to exercise what might be called *remote control* over other variables in the cluster by manipulating but one of them as an antecedent condition to the others (see fig. 7).

Artificial interlocking of variables. If the two lines are set up at different distances from the observer, as is done in the study of perceptual size constancy (Experiment C; see below, fig. 3), retinal projections are drastically unequal when bodily sizes are equal, and bodily sizes are drastically unequal when retinal projections are equal. This is represented along the scale of what will be called the manipulatory Variable (see § III/1) by the sharp separation of the POE's representative of the two experimental stimulus variables, *B* and *P*. Yet, possibilities for retinal projection of any given bodily length (such as particularly also that of the Standard) are narrowed to two constant alternatives (as long as the systematically predetermined distance ratio, in our example 1:6, is maintained), and vice versa for possibilities of physical object-correlates of any given retinal projection.

Although this type of co-variation, to be called artificial interlocking, thus defines a *"crucial experiment"*—rendering the old philosophical dilemma, namely, whether we "see" the retina or the outside world, "operationally" meaningful in terms of organismic functioning,—results have nonetheless to be considered contingent upon the rather specific, arbitrary choice of sizes, distances, and other conditions involved (see § VII/3).

Co-variation is here artificially set at less than 1, and this absence of a perfect correlation is in itself a definite step toward greater representativeness.

Artificial untying of variables. Whatever natural co-variation there is between two variables may be obliterated in an experimental setup so that their correlation is artificially reduced to zero within the special laboratory world created by the investigator. For example, identical ("constant") clothing and posture was requested of the soldier students serving as "social objects" in Experiment D in order to eliminate these factors as possible influences upon the subjects' judgments. Thus correlations between clothing habits, or general muscle tonus, on the one hand, and personality, on the other, which very likely would be found to exist outside the confines of the particular experiment, were destroyed in the experiment. Although the holding constant of a condition is a principal means of eliminating co-variation, there are other means to the same effect (see § VII/2).

b) *Representative co-variation.*—This is analogous to representative variation in that existing correlations are left undisturbed as they "normally" are. Correlations of 1 or of 0 as established under systematic experimental policies will, under these circumstances, be a rare exception rather than the rule. Therefore, these circumstances also confront the experimenter with a multivariate pattern of potential observation or evaluation. This latter pattern is automatically offered in statistical research with its inherent passivity and thus representativeness.

4. CLASSICAL-SYSTEMATIC AND REPRESENTATIVE DESIGN OF PSYCHOLOGICAL RESEARCH

The tradition of what might be called the *"classical" experiment* has been handed down to us from such famous origins in physics as Galileo's study of the fundamental laws of falling bodies. In terms of the distinctions introduced above, the classical experiment combines systematic policies of variation and co-variation—which in themselves may be taken to define *systematic design* in the wider sense of the word—with the rule of one variable. More specifically, the ideal formula for classical design may be summarized as follows: Have one independent variable and vary it systematically; hold all other conditions, at least those that may be relevant, constant; concomitant variation in the dependent variable will indicate the relationship in question. Do likewise with other variables and add up effects.

The crux of the psychological experiment is that among the variables potentially contributing to the final response only the ecological stimuli are characterized by an ease of control comparable to that attributed to all the variables in an experiment in physics proper. Many organismic variables, among them especially central conditions, are, for practical reasons, as a rule (1) open to little or no interference and (2) accessible only to indirect control of a kind that cannot be univocally scrutinized.

In order to approximate the classical ideal as nearly as possible, control of such conditions in psychological experiments is, in case (1)—important especially for *u*- or for *R*-variables,—usually replaced by the use of averages from within a representative distribution defined in terms of a broader class. Such averages may then be treated as rigid *quasi-constants* as long as the group remains identical, as occurs in the so-called "before and after" technique. To thus circumvent actual variability by computational after-the-fact elimination became especially important, since in the classical tradition the psychological experiment deals primarily with the "generalized human mind" (see Boring, 1929), excluding individual differences problems as much as possible.

In case (2)—usually combined with (1)—unknown factors inject an element of chance, with the result that control is merely more or less probable rather than rigidly univocal. An example of such *statistical remote control*, by antecedent condition, is the attempt to influence attitude, motivation, or other *a*-variables by the giving of an instruction or by the time elapsed since the last feeding (whereas such tests as amount of food eaten would use a subsequent condition). Similarly, *u*-variables may be held *statistically quasi-constant* by such devices as the "matched group" ("experimental" vs. "control") technique. The same may be achieved for *a*-variables—the importance of which, by the way, the classical psychologist tends to underestimate along with that of many external circumstances—by "balanced order of conditions" (a-b-b-a), whereby the experimenter gambles on the chance that uncontrolled systematic time effects such as fatigue or practice, especially when they may be assumed to be linear, will be neutralized along with the cancelation of random time fluctuations of attitude.

The emphasis generally placed in psychological experiments on the policy of "repetition" with large numbers of individuals or of trials—as contrasted with the singularity of the ideal "pure case" experiment in which all relevant conditions are known, see Lewin (1935, pp. 25 f.)—is a consequence of the comparative lack of manageability and accessibility of central organismic variables. It is in this vein that one of the founders of experimental psychology, Wundt (1907, pp. 307 f.; see also Woodworth, 1938, p. 2), has made repeatability a crucial criterion of the psychological experiment along with "active arrangement"—indicating that experiment creates new, rather than merely registers existing, fact—and planned "systematic variation."

It is predominantly by virtue of individual differences—long recognized as contributors to the response—that the entanglement of the psychological experiment with statistics has become an established fact in psychology (for a second historical inroad see § III/2). The modified *classical formula for psychological experimentation* may, then, be summarized as follows: All relevant external conditions (and there are supposedly not too many) to be systematically controlled, all internal conditions to be treated quasi-systematically by computational elimination of random variability.

However, the variation contained in such quasi-constants as the mean of a sample may be restituted in subsequent steps of added evaluation. When this becomes the main purpose, and especially when systematic stimulus variation is replaced by a constant stimulus pattern (as is especially clear in some of the "projective techniques"), the experiment changes into what is called a *mental test* (see § VI/2) as established by Galton in England and by Cattell in America. The independent variable then shifts from region *S*, the environment, to region *u*, individual differences.

Since representative sampling was thus permeating the entire domain of variables actually varied and investigated in psychological testing, it could no longer be ignored. It is this feature of representativeness—including the case that the entire population in question rather than just a sample is studied—by virtue of which differential psychology may be considered inherently "statistical."

By contrast, stimulus-response relationships are, with the limitations outlined above, capable of systematic treatment and have so far constituted the traditional domain of experimental psychology. This association developed primarily under the implicit presupposition that feasibility of systematic design is a unique chance that one must not let pass, wherever it offers itself. And although statisticians proper have tolerated and faced their difficulties cheerfully, and over and above this have developed a keen sense for the merits of representativeness where sampling rather than systematic control seemed unavoidable, all this was usually not done without at least tacit misgivings when the statistical approach was held up against the ideal of the classical experiment.

The task set for the subsequent sections of the present paper is to show that—quite aside from the asset inherent in the restitutability of individual differences in the shift from experiment to test—the deliberate replacement of systematic by representative design for stimulus variation and co-variation is the key to further progress in functional stimulus-response psychology. The goal of isolating variables, common to experiment and statistics, can then be achieved by partial correlation or other statistical devices; whereas, on the other hand, the policies of classical experimental design will prove to be fallacious with respect to this goal for reasons fundamental to the patterns of psychological functioning.

When restraining his interference with stimulus variables the experimenter actually duplicates, in a formal sense though with changed content, the patterns of research familiar in the study of individual differences. Thus, functional psychology is placed on a par with differential psychology with respect to basic methodological policy. Combining active command of the situation with representativeness rather than with artificial systematic design leads to the establishment of what may be called *representative experimental design*. Certain residuals of systematic procedure may hereby be retained to great advantage (see especially §§ V/2, VI/1).[4] Aside from representative variation and co-variation, representative experimental design also implies that the choice of the variables themselves should be sensitized to their biological relevance, as will be referred to later in this presentation.

The reorientation involved is not altogether novel as a factual policy, at least in such fields of stimulus-response research as social perception, in which one cannot help but invoke the ways of the statistician in certain respects. What was lacking heretofore all along the line, however—and with unfortunate consequences, as will be seen—is the explicit conceptual recognition of stimulus representativeness as a respectable universal research principle.

III. The Classical Psychophysical Experiment and Its Relationship to Error Statistics

1. EXPERIMENTAL DESIGN, TYPE A. EXAMPLE: GALTON BAR

In our model sequence of experiments in perception, the classical experiment is exemplified by Experiment A, Length Discrimination on the Galton Bar.[5]

The experiment may be visualized by means of figure 1 and figure 4, *A;* a more abstract summary of the experimental design is given in figure 7. The independent variable is said to be "length," or, more precisely, the "length difference" between a constant "Standard" and a "Variable"[6] as given by the two (near-)halves of the bar.

[4] Thus, the treatment of traditionally experimental stimulus-response problems will absorb more and more of the general outlook of statistics. Nevertheless, it should be noted that certain differential psychological problems have recently benefited from the adoption of features of the experimental approach; to cite only one example from the psychological study of heredity (the pioneer field of statistics), we refer to the systematic breeding of "bright" and "dull" strains of rats by Tryon (1940).

[5] See Galton's works as described by Pearson (1924, Vol. II, pp. 221 ff., referring to a report on Galton's Anthropometric Laboratory published in 1885); see also Titchener (1901/1905), and Garrett (1941). Data on visual length discrimination were published by Weber as early as 1846, reporting threshold values of 1/50 or even 1/100 for lines; his classical psychophysics of touch is even older (see Boring, 1929).

[6] This term, when spelled with a capital, may further be specified by adding the attribute "manipulatory" in order to avoid confusion with the term in the more customary sense of "experimental variable," variate, dimension, denomination, or "treatment" (as in the analysis of variance). As can be seen from figures 2, 3, and 4, Standard and Variable participate in a common technical denomination, usually that of one of the experimental variables (in our case, "bodily length"); whereas various points along the manipulatory Variable scale may come to represent different variables (dimensions), both of the stimulus and the response kind.

All other relevant conditions are said to be held constant; this is notably true for the spatial orientation and the distance from the observer of all lengths entering the comparison, as well as their surroundings if a homogeneous background is used. However, especially in the earlier phases of psychology, only relatively few variables, such as those listed, were suspected as possible contributing factors to the responses,[7] even fewer than had to be recognized in some of the classical physical experiments that served as a methodological prototype.

Fig. 1. (Among other sources see Garrett, 1941, fig. 47, A.)

Experiment A. Galton Bar.—A good example of the emphasis on precision apparatus, with neglect of genuinely functional problems, in an era of psychology sometimes criticized as given to "brass instruments." In the best classical tradition, the two horizontal halves are presented against a homogeneous background.

The dependent variable is constituted by the responses given by the subject. Results may be summarized in a score giving the "difference threshold" for the task in question.

This threshold is defined either as a "difference limen" (DL), in turn based either upon the relative number (with the "method of constant stimuli") or upon the relative number and certain features of the location (with the "method of limits") of uncertainties of judgment. Another measure of the difference threshold is the standard deviation (SD) of the points of subjective equality (PSE) of repeated presentations. These in turn are defined either in terms of the transition points of subjective judgment (with the method of constant stimuli) or as active settings to apparent equality by individuals ("adjustment method"). With the first-named two of the three classical psychophysical methods listed, the manipulatory Variable is varied by the experimenter systematically, i.e., in a predetermined series of presentations along a scale of equal steps. Threshold values differ, depending on the method used. See Woodworth (1938, chap. xvii).

Although classical "psychophysics" is primarily concerned with sensory qualities and their intensity and other attributes—as, for example, with color sensations and their brightness or with tonal sensations and their pitch and loudness,—all its essential methodological features are present in our case as well. The entire approach is an outgrowth of a type of

[7] Thus, the frequently reproduced illustration shown in figure 1 gives no inkling of a covering up of the potentially disturbing mechanical accessories involved, and captions usually do not mention the desirability—existing, at least, from the classical point of view—of a homogeneous background. If at all, classical experimental psychology is inclined to establish clearly the influence of neighboring variables only so far as the *immediate* spatial or temporal surroundings are concerned, such as in the case of ("marginal") color contrast, after-images, and the like.

introspective analysis, known by the somewhat misleading term "Structuralism," which dominated the psychology of the nineteenth century and was chiefly represented by Wundt in Germany and by Titchener in America.

The intrinsic shortcomings of the classical psychophysical design, especially what will be called its "clustervariate" rather than truly univariate character, will appear in the course of our further presentation.

2. EARLY STIMULUS-RESPONSE STATISTICS: ERRORS OF MEASUREMENT AND OF RELATED TYPES OF OBSERVATION

The invasion of the pure experiment by the statistics of individual differences and the positive use made of individual differences in testing as described above (§ II/4) are not the only ways in which traditional psychology has come in contact with statistics. A third is given by an elementary kind of stimulus-response statistics, of the type described above in connection with the listing of the various measures of the threshold and climaxing in such generalizations as the Weber-Fechner law.

Considerable effort on the part of psychophysicists working in the wake of the classical tradition (such as Fullerton and Cattell, 1892, and Cattell, 1893) has gone into the treatment of perceptual responses in terms of observational errors of the kind known in physical "measurement" and dealt with in error statistics ("normal law of error"). It is to be noted that this is statistics of variability—rather than correlation statistics as is the type of representative stimulus-response research suggested in the present paper.

Indeed, the experiment with the Galton bar is, in its spirit, as close to measurement as any psychological experiment conceivably could be; comparison of lines has been made but little more difficult by placing them side by side, rather than parallel and as closely "coincident" as possible as in measurement proper. In such cases as "photometry" physical observation and psychophysics become altogether indistinguishable since we are then dealing with the same type of "response" that is used in establishing a "stimulus." Experiments of this kind actually belong to both disciplines at the same time. Thus, if one should wish to extrapolate the series of our model experiments backward beyond (A), "measurement of length" with its observation of spatio-temporal point coincidences of the two ends of a line with marks on a scale would be the logical choice.

In the course of time, and right up to the immediate present, psychologists have thus become firmly accustomed to think of the role of statistics in psychology in a threefold way:

First, and negatively, in an experiment, statistics comes in as a disturbance through individual differences, and is to be eliminated as painlessly and as much without leaving a trace as possible.

Secondly, and positively, the same individual differences become the backbone of one of the most important fields of psychology—mental testing. It is this domain with which the concept of psychological statistics is usually associated and which has in turn contributed appreciably to the development of such mathematical statistical instruments as the correlation coefficient.

And thirdly, on a siding, and chiefly back in forgotten times, there is the error statistics of classical psychophysics, seemingly a dead end so far as psychology is concerned, and continuing into our times in a mere trickle of frozen tradition.

The broadening of psychophysics into a multivariate and functionally representative discipline—a discipline which thus is not only experimental but is at the same time statistical—will be the topic of the subsequent sections (§§ IV to VII).

IV. Proximal Multidimensional Psychophysics of "Gestalt" Problems: Intraorganismic Field Dynamics

The stress in a second phase is on interaction of variables in a broader context. This is the stage of systematic "multidimensional" psychophysics which forms a transition from the classical to the fully representative stimulus-response experiment. Two steps may be distinguished here, to be discussed in the present and the next section.

1. EXPERIMENTAL DESIGN, TYPE B. EXAMPLE: MÜLLER-LYER ILLUSION

The first of these steps, inaugurated by the early Gestaltists of the Graz school during the last decade of the nineteenth century, dealt predominantly with the negative, disruptive aspects of interaction in a field, that is, with "inadequacies" (Benussi) of perceptual responses in comparison with the actual stimulus. Examples are the geometrical optical illusions, or apparent movement.

From here, the Berlin school of Gestalt psychology proper (Wertheimer, Köhler, Koffka) expanded in the second and third decade of the present century into a general dynamics of form perception. As the present writer sees it, this development climaxes in the "law of Prägnanz," a principle still confined, as were the illusions, to an emphasis on a type of interaction within a broad field which is in effect upsetting rather than establishing correct, or "univocal," correspondence between environmental variables and organismic responses.

Our example for this phase of development of the psychology of perception is Experiment B, an illusion described by F. C. Müller-Lyer (1889). In figure 2, otherwise unpublished results of a recent thesis by Marianne Müller (1935) are shown together with the illusion pattern itself. Again the general design of experiments of this type is summarized in figure 7.

The "crucial" character (see § II/3, *a*) of this experiment is established by the fact that the artificial tie set up between the variable "length," *L*, and the variable "adjacent, or 'fringing,' area," *F*, has been replaced by artificial interlocking, giving the two stimulus variables a measure of independent variability in what is in a formal way analogous to the type of "multidimensional" psychophysical experiment as it is known from recent research in more purely sensory domains (see below, footnote 11). More specifically, it is one of its major *neighboring* variables that length is "confronted with" (see Brunswik, 1934, § 28).

As compared with the classical psychophysical experiment, experiments of such design are capable of revealing the intensive "confluxion" that exists among certain neighboring variables, making them conjointly highly contributing to the response. In our particular instance, fringing area proves functionally even more powerful than length (see fig. 2 and below), and this in spite of the fact that it is length which is apt to be singled out by the classical experimenter as the independent variable, both by virtue of its introspective prominence as what may be called the "figural" variable and by virtue of its being the most conveniently manipulable or measurable aspect of the situation.

Ecological variation of a systematic type, and, even more, spotty canvassing that may be considered an implicit and rather clumsy attempt at controlled sampling of problems, is comparatively abundant in the investigation of illusions of this kind. The Müller-Lyer illusion inaugurates what Boring (1929, p. 630) has characterized as the "decade of the optical illusion" in the history

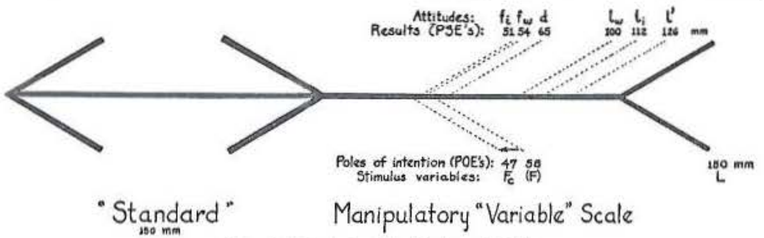


Fig. 2. (Data by M. Müller, 1935.)

Experiment B. Müller-Lyer Illusion.—Main pattern of two horizontals, with by-pattern of three pairs of radiating lines (arrow tips) pointing in opposite directions. The angles in this particular version are 30°. Attitudes are designated as l or f depending on directedness of the subject toward length of horizontal lines, or toward fringing areas, respectively, whereas d stands for a "diffuse" (undirected, aesthetic) attitude toward "good balance" of the two halves of the drawing. Subscripts w and i further specify whether the directedness was natural, relaxed, "whole-perceiving," or else analytic, "part-isolating." Both these types of attitudes stay within perception proper, although there is more concentrated effort in the case of i. Rational control through abstract knowledge or thinking is here only applied to length, in a "critical attitude," l'. Results are averages of 16 subjects, based on the adjustment method, with a mixed order of attitudes. The fringing area is at first tentatively defined as the area cut off by imaginary straight lines connecting the open ends of the arrow tips. Thus the value for (F) corresponds to the open distance at the mouth of the enclosures adjacent to the Standard. Results of additional experiments by Miss Müller with actually closed figures make a correction, indicated by the small arrow leading to F_c, seem appropriate. This correction represents an outward bulging of the areas which must be assumed to "belong" perceptually to the horizontals and thus to be responsible for the illusion; the greater bulge is on the more open, right-hand side.

of psychology, covering the 1890's. A considerable array of illusions of size, direction, and curvature were described, including several variants of the Müller-Lyer pattern. Soon, analogous instances in the tactual field were discovered. Furthermore, variation of such features as the angle between horizontals and radiating lines in the Müller-Lyer pattern was studied by Brentano and by Heymans.[8] Marianne Müller showed that in unconstrained length comparison there is an almost linear increase in the amount of illusion as this angle decreases from 60° through 45° and 30° to 15°.

The range of ecological variables contributing to judgment of magnitude and of number goes so far as to include such socially conditioned and emotionally or motivationally loaded factors as the monetary value of coins; see Brunswik (1934, reporting unpublished studies by Fazil and Zuk-Kardos), Ansbacher (1937), and Bruner and Goodman (1947).

[8] The history and theories of the illusions may be found in Boring (1942, pp. 238–246, 252 ff., 261); summary in Woodworth (1938, pp. 643 ff.). It should be added that the above-suggested treatment of the Müller-Lyer illusion as a length-area "assimilation" ("dissimilation," or contrast, is the opposite alternative, both terms being more customary than their common heading, confluxion), being merely one of a series of synonymous or "tautological" descriptions of the objective situation, does not prejudice in terms of any of the numerous physiological "explanations" listed in these two sourcebooks.

2. ATTITUDE AS STUDIED IN MULTIDIMENSIONAL EXPERIMENTS

By the same token, the great importance of the subjects' attitude (*Einstellung*), in the specific sense of directed attention toward a certain variable, became evident. A study on the Müller-Lyer illusion comparing the normal whole-perceiving with a part-isolating attitude was published by Benussi in 1904 (see Woodworth, 1938, p. 649). The recent data of Marianne Müller demonstrate the great flexibility of the responses when attitude shifts from one of her no less than six varieties (see the legend to figure 2) to another.

The traditional measure of illusion is the *constant error*, CE = Mean PSE − Standard (or POE). It is occasionally called "systematic error" since there is little that is constant about this error. On the contrary, variability of judgment—both intra- and intersubjective—is usually increased radically in comparison with the classical experiment in which the response is sheltered against disturbances. Since length-area confluxion is an instance of "perceptual compromise," the constancy ratio, c, to be introduced with Experiment C, may be used as an alternative; the two POE's, L and F, constitute the poles of the scale, 100 and 0, respectively. (Thereby, capital letters may be used to designate either the stimulus variable or the POE representing it on the manipulatory Variable scale, as lower-case letters are used for attitude as well as for the corresponding response variables and concrete PSE's obtained in taking those attitudes, so long as there is little danger of misunderstandings.)

With the particular version of the Müller-Lyer pattern presented in figure 2, Marianne Müller obtained a CE of 50 mm, or 33 per cent of the Standard, in the case of natural, unconstrained attitude toward length, l_w. This is somewhat more, perhaps, as one may find in a class demonstration, and possibly to be accounted for by the alternation with other attitudes emphasizing area. (In fig. 2 the illusion is probably somewhat enhanced to the reader by the "subdividing" dotted lines indicating the results.)

This error diminishes by shifting to the length-isolating attitude, l_i; it is approximately cut in half by superimposition of intellectual self-criticism (l'), a result quite typical for cases in which the subject remains "without knowledge" of the correct solution, as was the case here (see also Experiment C). In group experiments in the classroom, critical attitude "with knowledge" was found capable, on the average under certain conditions, of overcoming the illusion completely, although large SD's obtained point toward a good deal of conflict in such situations.

The area pole, on the other hand, is well attained in the respective attitudes f_w and f_i, even without resort to a critical attitude (see also legend to figure 2). The functional dominance of F over L is further revealed by the fact that responses in the diffuse attitude, d, cling much more closely to the former than to the latter "pole of intention."

Within the scope of the a-variables, encompassing all intra-individual differences, lie what are known as practice effects. Like analytic attitude, practice—in the sense of sheer repetition—was found to reduce the illusion almost to zero, although the effect does not always stand up to reversal of the figure (see Woodworth, 1938, pp. 195 f.) and is thus not highly generalizable.

As to inter-individual differences (u-variables): Thurstone (1944, pp. 63 f., 89 f., 95, 120 f.) found that among his close to 200 subjects, mostly students, women were slightly more susceptible (less resistant, $r = -.14$) to the Müller-Lyer illusion than men.[9] In 76 students, Ehrenstein (1935, pp. 58 f.) found averages of a variety of illusions to be 25 per cent higher in those oriented toward the humanities than in natural scientists and those who in school had excelled in mathematics and physics and thus could be regarded as analytically inclined.

[9] The over-all average for the illusion computed from his table on page 64 is 22 per cent, considerably less than the 33 per cent computed from our figure 2. Since Thurstone uses a 45° angle instead of the 30° angle on which the latter value is based, a discrepancy in the direction obtained is in line with the above-quoted findings about the effect of the angle upon the illusion.

Thurstone further reports that correlations with the "primary mental ability" of reasoning are close to zero although the perceptual "factor" tying together the various optical illusions correlates .14 with Reasoning, indicating less illusion in the more intelligent, a fact anticipated by Crosland, Taylor, and Newsom (1927). Binet (1895), and later Pintner and Anderson (1916), found children to be slightly more susceptible to the Müller-Lyer illusion than adults (see also below, § V/1).

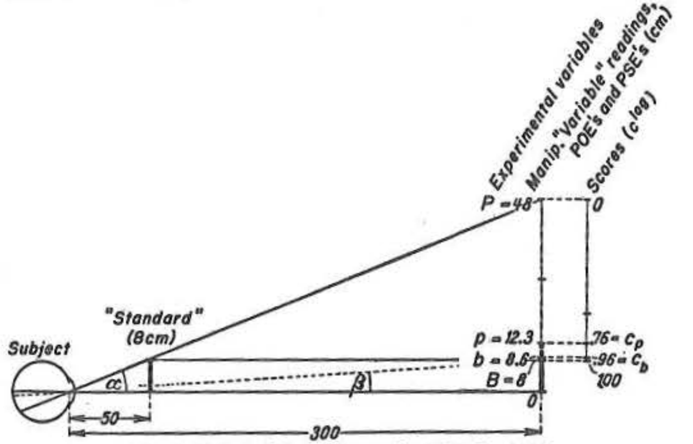


Fig. 3. (Adapted from Brunswik, 1935, fig. 111.)

Experiment C. Perceptual Size Constancy.—In actual procedure, the head of the subject will be placed so as to avoid overlapping of Standard and Variable which would, especially if nearly complete, greatly facilitate photographic perception. Results shown are fictitious, but not atypical, though perhaps on the higher side as to c when geometric objects not associated with any particular familiar size are shown in an indoor setting that furnishes a normal array of distance cues (see text). Unannotated marks about the middle of the Variable- and c-scales indicate the midpoints, 28 cm and c = 50, respectively. The difference in their location indicates the downward shift of the sensitivity of the scale in the direction of the smaller magnitudes as the result of the use of logarithms instead of numerical values.

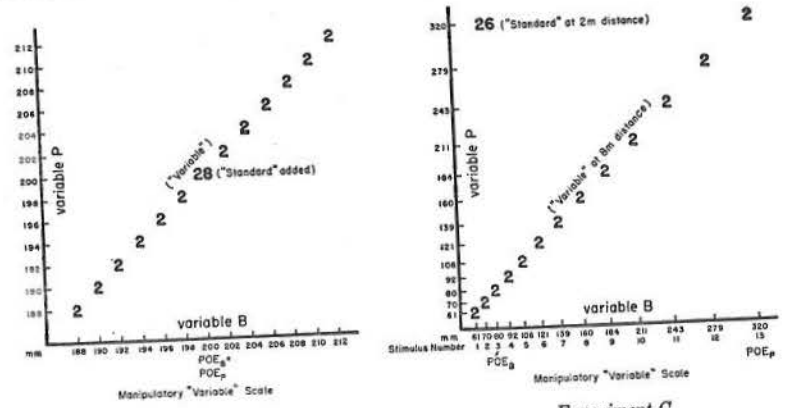
V. Distal Multidimensional Psychophysics of the "Thing-Constancies": Stabilization of Relations with the Remote Environment

1. Experimental Design, Type C. Example: Perceptual Size Constancy

The second step in the development of multidimensional psychophysics differs from the first in that it is crucial with respect to an independent variable which is confronted with one of its *mediating* conditions rather than with one of its neighboring variables as in the Müller-Lyer experiment. In terms of experimental design this means that the artificial tie of distal and proximal variables familiar from Experiment A, Galton bar, is replaced by artificial interlocking. It is this specifying reaching out into the environment which renders this approach "functionalistic" in a way demonstrated more clearly in figure 7 (see also fig. 10).

In the size-constancy experiment chosen to represent this step (Experiment C, fig. 3)[10] retinal extensions are pitted against the sizes of objects in the

[10] History of the problem by Boring (1942; concluding sections of chaps. vii and viii, pp. 255 f., 262, 288–303, 308–311). See also Fröbes (1923, pp. 317 ff.; 1935, pp. 25 f.). Concerning the 1910's as the beginning of full maturity of the research in the thing-constancies, see below, subsection 2.



Experiment A

Scattergram for the Galton bar, envisaging a Standard of 200 mm and 13 values of the Variable, at a common distance of, say, 2 m. Owing to the prearranged equality of distance on the two sides, influences of retinal size, *per se*, must always be proportional to the bodily length of the lines. Such influences are also equal on both sides if the lines themselves are equal. The "channeling of mediation" underlying this state of affairs is reflected in the diagram by the identical location of the points of objective equality (POE) for *B* and for *P* along the manipulatory Variable scale. Owing to the resulting lack of disturbance and absence of ambiguity, accuracy of perception is high, and step intervals may correspondingly be small; in further consequence, differences between geometric and arithmetic progressions are negligible, and the more convenient arithmetic gradation may therefore be chosen, in accordance with the policy prevailing as a matter of course in nearly all classical psychophysical experiments. Step intervals are here 2 mm, i.e., one one-hundredth of the Standard, arranged symmetrically about the common POE. Further assuming that there are two presentations for each value of the Variable, say, in a pair of ascending and descending series, accompanied by a showing of the Standard for simultaneous comparison in each of the 26 presentations, the scattergram shown above is obtained. All size exposures within the framework of the experiment fall in a diagonal, establishing perfect correlation (r = 1.00) between the variables *B* and *P*.

Experiment C

Scattergram for size constancy, assuming a Standard of 80 mm at 2 m distance, and a Variable at 8 m with geometric (logarithmic) step intervals allowing for ten spaces between the two poles of the design, $POE_B = 80$, and $POE_P = 320$. The constant logarithmic increment is then .06021, resulting from a constant ratio of the geometric series of stimuli of $\sqrt[10]{4} = 1.149$, corresponding to an increase of about 15 per cent over each preceding value; the 4 under the root represents the distance ratio, 320/80, and thus the ratio of the poles. Two extra steps beyond POE_B are allowed to provide for the frequent cases of "over-constancy," i.e., overcompensation for disruption by distance, bringing to 13 the number of stimuli along the Variable scale. Following otherwise the procedure outlined at the left, a correlation coefficient between *B* and *P* (hardly appropriate in view of the bizarre scattergram in this systematic design) has been computed as an oddity, using the stimulus numbers rather than the numerical values. The artificial relationship is found to have been thrown off by the isolated cluster of 26 objects in the upper left-hand corner from the value of 1.00 for the Galton bar to a value of .08 for the present arbitrary size-constancy situation. This may also serve as an illustration of the possibility of a complete disruption—by the technique of artificially interlocked co-variation rather than by the more customary process of the holding constant of a variable—of whatever natural correlation there may exist between *B* and *P* (see § VII/2, 3).

Fig. 4.

Examples of Scatter Diagrams for Systematic Univariate (more Correctly, Cluster-Variate) *and Multivariate Experimental Design as Represented by Experiments A and C.*—By emphasizing adjacent area rather than retinal projection as the second variable, a similar comparison could be made between Galton bar and Müller-Lyer illusion.

environment. It separates the variable "bodily length," in this context best to be named B, from their direct retinal projections, the "physiological" ("photographic") sizes, P. Thus the "channeled mediation" (see § II/3) defining the artificial tie between B and P characteristic of the first two experiments is given up. The resulting change from perfect to imperfect correlation between the two stimulus variables B and P that makes possible an isolated treatment of the two variables may best be visualized by comparing the two scattergrams shown in figure 4.

Experiments A and B are both, at least in the classical interpretation, to be considered proximal in character (by rigid remote control of retinal conditions through handling of distal manipulables); they are certainly no better than proximo-distally neutral, or inconclusive. Experiment C, on the other hand, is capable of differentiating between distal vs. proximal determination of responses. Both Experiments B and C are multidimensional.[11] In retrospect, the alleged unidimensionality of the classical Experiment A must appear fallacious. Neither of the variables B or P is really isolated. They are merely inseparably tied by what would better be called a "tied-variate" or "cluster-variate" rather than a univariate design; hence, even if they should both be accurately responded to when tied, there is no saying which one of the two is the specific stimulus or whether the response would be a function of both of them at the same time when they are untied. It is obviously in the nature of things that not all variables except one can be held constant, if only the scope of experimental observation is extended beyond the immediate conditions to include past and remote causal variables. This latter must be done if the goal is an inventory of the gross efficiency of organismic functioning against the background of disturbing factors rather than a study of the "fundamental" minutiae of organismic technology.

The major result is given by the apparently paradoxical approximate invariance, or "constancy,"[12] of the appearance of physical bodies, despite drastic changes in distance and thus in retinal representation (for specific values see below). In the statistical sense defined above (§ I/3), contributingness of the mediating proximal variable, P, is thus found to be at a minimum; this is an important instance of causation without correlation. The final response is to a high degree "focused" on the distal variable, B. Their relationship may thus be called "stabilized"; or the variable B may be called nearly "attained" (Brunswik, 1934) by the response variable b.

[11] Experiments so far reported in the literature under the heading of "multidimensional psychophysics," e.g., those by Richardson (1938), or Stevens' well-known analysis of tonal attributes (see Boring, 1942; Woodworth, 1938, p. 509), are proximal in character. In fact, they even deal with sensation in the narrower sense—the subject matter of classical psychophysics proper,—whereas the Müller-Lyer illusion deals with figural elements such as apparent length; the latter likewise ignores, however, the distal-proximal alternative.

[12] This use of the term in "thing-constancy" has nothing to do with its previously mentioned use in "method of constant stimuli" (often abbreviated to "constant method"); still another concept, that of the "constancy hypothesis," refers to the assumption of a one-to-one relationship between sensation and the elements of *proximal* stimulation which, according to an allegation by Gestalt psychology, is implied in classical structuralism. Such constancy, if true, would in effect contradict the intuitive character of the perceptual thing-constancies in which the relationship to the *distal* stimuli appears stabilized. (Needless to add that the term "constant error" refers to a still further usage.)

On the average, existing errors point in the direction of what the present writer has suggested calling a perceptual "compromise" effect involving bodily and retinal size (see Brunswik, 1928, 1934; Thouless, 1931). This principle and its generality will be further discussed below (§ VII/3).

Unless a subject is specifically trained in perspective drawing or painting, even a change of attitude toward an intent to perceive in proportion with conditions existing on the retina of the eye (or on a real or imagined projection screen) is not likely to obliterate perceptual size constancy to any drastic extent.

The degree of perfection of perceptual size constancy may, especially in systematic experiments, be measured by the *constancy ratio*, suggested by the present writer (1928) for the analogous case of color constancy. In adapted notation, it is

$$c_{BP} = \frac{P - r}{P - B}$$

whereby r designates responses (PSE's) obtained in any of several possible attitudes.[13] This ratio measures the completeness of the functional "transformation" of the relatively large retinal image of the Standard, subtending the visual angle α, to the proportions of that of the far Variable and especially to β, defining the visual angle under which the bodily equivalent of the Standard is projected upon the retina. It does so by giving the degree of approximation of r to B as a fraction of the ideal transformation given by the full distance between the "poles" B (1.00, represented by the first subscript to c) and P (0, second subscript). Small-letter subscripts may be added in parentheses, or may be used alone whenever there is no danger of misunderstanding, to specify the attitude in which the response has been given. The poles themselves represent perfect size constancy (1.00) and complete lack of transformation or perfect retinal attainment (0), respectively. Values of c of more than 1.00, indicating overcompensation for the difference in distance ("overconstancy"), are of course entirely within the scope of the formula; in fact, they are quite frequent in individual cases or under atypical (nonrepresentative) situational conditions. It must be kept in mind, however, that they represent deviations from the correct response just as much as do values under 1.00. Admitting these values as computational expedients helps the ascertainment of the average trend in the results which may point to a value quite close to perfect thing constancy, with a good proportion of the scatter falling beyond, rather than short of, the ideal achievement.

(All this seems to fit best a type of instruction asking for r to be given in the subsequently discussed natural attitude, b, which is directed toward the distal stimulus variable, B, thus specifying the above-given constancy ratio to $c_{BP(b)}$, or, in brief, c_b. In attitude p, directed toward the proximal stimulus variable P, on the other hand, not B but P—more precisely, not POE_B but POE_P, see IV/2—is the correct answer. One may, therefore, feel inclined to have B and P exchange places in the formula, thus defining an inverted constancy ratio

[13] In translating the notations from the original article (Brunswik, 1928), Woodworth (1938, p. 605) has chosen S (for "stimulus") as the analogue to our present P, A (for "albedo," the reflecting power of a surface, p. 598) as the analogue to our B, and R for our r, in what he has labeled the "Brunswik ratio" (BR, as in Cohen and Quinn, 1946, where a short computation method is described).

For reasons of the Weber Law and other, even more intrinsic, reasons (see Brunswik, 1933, 1934), a substitution of the logarithms of B, P, and r for the numerical values (first explicitly suggested by Thouless, 1931) is to be preferred, especially when the "span" between the poles is large, as is usually true in experiments on size constancy. Actually, the same effect as by the use of logarithms can also be achieved by varying the manipulatory Variable in geometric (logarithmic) intervals rather than in the customary arithmetic series. The scale of gray papers used in the original article of the present writer constituted such a geometrical progression of stimuli. The order numbers assigned to each of these stimuli (as also seen in the right part of fig. 4), when used to replace the actual magnitudes along the Variable scale in the numerical version of the formula, lead directly to what may be specified as a c^{log} in contradistinction to a simple c^{num}.

$c_{PB(p)}$ in which P coincides with 1.00 and B with zero. However, such a dual procedure has proved too confusing in practice; the abbreviated symbol c_p used below therefore refers to a noninverted $c_{BP(p)}$.)

Instructions to the subjects constitute a special problem in all experiments on thing constancy. To counteract "stimulus errors," i.e., intellectual interference that may result from knowledge in geometrical optics or in physiology or from specific information, the casualness of the everyday, practical, "naïve-realistic" attitude b has to be drastically emphasized by analogies referring to such features as the tallness of people in feet and inches, or the size of cars and of other objects of utility varying within narrow ranges, to be established at varying distances. On the other hand, reference to perspective drawing or to measuring on a photograph will help to establish attitude p. An abbreviated set of instructions for these two properly perceptual attitudes, as well as the corresponding two "critical" attitudes inviting the stimulus error by letting the person imagine that he has to take an actual "bet" on the judgment, is given in Brunswik (1944, pp. 4 f.).

Holaday (1933) studied size constancy under purposely somewhat less favorable ecological conditions than those assumed for the fictitious data shown in figure 3. Cubes were placed on the empty floor of a large auditorium which was otherwise left unchanged, with the subjects' heads but slightly raised above the floor. Results obtained are here recomputed, using the logarithmic instead of the numerical values of the original data, and averaging the data for the four "distance constellations" of Standard and Variable (1 : 2, 1 : 4, 1 : 8, 2 : 8 m) and for the ten subjects used. They are as follows: .86 for c_b, representing the most natural, unconstrained attitude, in contrast to .39 for c_p. Corresponding critical attitudes yield .94 for $c_{b'}$ and .25 for $c_{p'}$, cutting errors in both directions approximately in half. Group experiments conducted by the present writer in the laboratory, using flat triangles in a manner described in Brunswik (1935, p. 87), with distance constellations ranging from 2 : 8 to 6 : 12 m, yield on the average a $c^{\log}$ of about .9 for b, of about .6 for p, and of about .2 for p'. Experimental series were started at the nonintended pole in each case, loading the procedure against the hypothesis of good "attainment" of the intended pole by the action of what is known as "central tendency of judgment" (see Woodworth, 1938, pp. 445 ff.).

A similar increase in the "shift span"—but this time within the perceptual system proper—has been found to result from practice ("without knowledge" of correct responses) in experiments by Klimpfinger (1933*a*; see also Brunswik, 1934, fig. 7) on the related problem of the "shape constancy" of ellipses rotated into the third dimension. Constancy ratios rose from about .6 to about .9 for b, and fell from about .3 to about .2 for p (likewise an improvement), in the course of twenty-six training series scattered over a period of one month.

Development of color- as well as size- and shape-constancy has been found to have reached a considerable degree at the ages of two or three years, with a definite climax from about ten to about fifteen and a slight decline in adulthood (see Brunswik, 1930). This decline is probably to be interpreted as a shifting toward more abstract processes or functions, such as memory and thinking, "analogous" (Werner, 1940) in furnishing orientation in the environment to the relatively autonomous, more primitively organized perceptual system proper. A similar increase up to the age of ten and a decline from then on was recently observed for a certain form of the vertical illusion by Würsten (1946). Concerning the first intimation of the constancy mechanism at the age of about six months see Brunswik and Cruikshank (1937) and Cruikshank (1941). Some of the remarkably high constancy ratios found in animals are reported by Woodworth (1938, chap. xxiv).

Retest reliability, after two years, has been found as high as .92 for an experiment in shape constancy (Thouless, 1938). Different tasks in the same special branch of the constancies were found to correlate from .3 to .9 by Sheehan (1938), using 25 subjects only, whereas Thurstone (1944), with his close to 200 subjects, found intercorrelations to be only .12 or under. Intercorrelations among size-, shape-, and color-constancy, throwing light upon the problem of test validity and of the trait character of the perceptual constancies, have been reported by Sheehan (1938) and by Thouless (1932, 1936, 1938) as between .3 and .6. With 76 subjects, Weber (1939) found a higher degree of size constancy in extraverts than in introverts. Thurstone (1944, p. 95) found correlations of brightness-, size-, and shape-

constancy with the factor of Reasoning to be practically zero; women are slightly superior to men in the tests involving brightness constancy (r = .21) and size constancy (r = .17), a finding which in view of the large number of subjects should not be ignored. (The treatment of the constancy problem by Thurstone suffers from lack of consistency with respect to whether establishment of high constancy—good "object judgment"—or the ability to circumvent this perceptual mechanism—good "sensory judgment"—is taken to define the perfect score. For brightness the former, for shape the latter policy is being followed. For size, indications seemed ambiguous to the present writer; see pp. 70 f., 89 ff.)

As to the variation of the distance cues available see subsection 2. A purely ecological dimension systematically varied by Holaday is distance constellation. For the combinations given above, augmented by others involving the distance of ½ meter, there is, under the indoor conditions described, a steady and significant decrease of $c^{\log}$ with increasing absolute distances when distance ratio is held constant, from .90 for ½ : 1 m to .71 for 4 : 8 m. Recent experiments by Gibson and Glaser (Gibson, 1947, chap. ix) with outdoor distances up to nearly 800 yards show almost undiminished accuracy in terms of "constant error." Since what corresponds to the manipulatory Variable is at a constant distance of 14 yards, this result means an increase in c with increasing distance discrepancy.

Attempts to fit simple generalizable curves to data of this kind would probably not be rewarding in view of the large number of ecological factors contributing to the establishment of thing constancy, although Hsia (1943) has recently made an attempt to formulate such a law for the degree of color constancy as a mathematical function of the degree of discrepancy of illumination between Standard and Variable.

Stabilization effects of the kind studied in the thing constancies are of the very essence of life, and their recognition is one of the most fundamental tasks of a functionalistic psychology concentrating on questions of utility and adjustment. They appear paradoxical only as long as isolated mediating stimulus variables, such as retinal size, is considered apart from other conditions. Actually, all thing constancies require "multiple mediation," i.e., the contributingness to the response of proximal "cues" representing "circumstances" of the situation viewed, in our case the distances from the observer of the objects perceived (Bühler's "duplicity principle," 1922; see also Brunswik and Kardos, 1929). This is further illustrated below in figure 10 and the accompanying discussion.

An example of a physiological stabilization mechanism is "homeostasis," the constant maintenance of conditions in the internal environment, such as blood temperature (see Cannon, 1932). Naturally grown or artificial lenses capable of focusing in the literal sense of the term present the principle in a nutshell. Stabilization mechanisms of an artificial kind are found in bombsights and gun stabilizers of tanks. (For further discussion see Brunswik, 1952.)

Fundamentally, stabilization mechanisms must be considered as special cases of confluxion, and thus in the last analysis akin to the field dynamics studied by Gestalt psychology. The functionalism reflected in the study of perceptual thing constancy, however, acknowledges a positive, rather than the negative (organism- rather than stimulus-centered) aspect of these dynamics. Stabilized relationships with the environment are biologically useful adjustments, especially when they anchor organismic orientation to properties of more or less remote, more or less vitally relevant solid objects of potential manipulation and locomotion such as landmarks, tools, enemies, or prey which themselves are usually fairly stable or predictable (see also § VI/4).

Confluxion and especially focalization of such vast proportion—and especially such vast consistency—as are encountered by psychologists have not been faced by the classical physicists. Vice versa, experiments patterned after those of classical physics, though feasible in psychology (as shown by the classical psychophysical experiment), bypass the vital type of information given by multidimensional stimulus-response research, especially by the functionalistic type separating the distal from the proximal variables.

To give only one example: threshold values on the Galton bar are, under certain conditions, about 1/40; a person displaying a threshold twice or three times as large under similar conditions would be quite an extreme case. Yet, a person seriously deficient in the mechanism of size constancy may easily be off a thousand times the value of the threshold, e.g., if he would—as Helmholtz reports in one of his early childhood memories—mistake for dolls people walking along the high balcony of a church steeple. Thurstone (1944) found correlations of size-, shape-, and color-constancy with the classical differential threshold (established after the reaction time method) for size of circular areas—a close equivalent to our Galton bar experiment and the only pertinent classical psychophysical experiment in his battery of perception tests—to be −.06, −.04, and .03, respectively. Thresholds for magnitude can therefore by no stretch of imagination be construed to be a valid test for thing constancy.

Furthermore, visual acuity proper, although undoubtedly highly instrumental in the utilization of "minimal cues" of distance, is likewise in itself in no way a direct measure of the confluxion and stabilization mechanisms involved in size constancy. An indication of a certain amount of correlation may be derived from the material presented in table 5 of the study by Holaday (1933). For the ten subjects used, the rank-order coefficient is .57. Owing to the small number of subjects, however, significance is not established (p is no better than .075).

2. THE TECHNIQUES OF SUCCESSIVE ACCUMULATION AND OF SUCCESSIVE OMISSION OF FACTORS

The size-constancy experiment can thus be considered a more representative type of experiment by virtue of the mere fact that it checks on stabilization, demonstrating at least the possibility, in certain cases, of distal focusing of perception. But otherwise the schema given in figures 3 and 7 does not reveal departure from systematic design. There actually is such a departure, however, at least in the way in which such experiments have been conducted more recently.

The topic itself has been known to, and discussed by, such classicists as Helmholtz (1866) and Hering (1879); and a real start on experimentation was made by Martius as early as 1889, about the time research on the geometrical-optical illusions was established (for references to the history of the problem see footnote 10, above). But throughout the first two decades all but one or a few cues—usually the classical favorites in the family of perceptual distance criteria, binocular disparity and convergence—were experimentally eliminated or minimized by the use of such paraphernalia as screens, a darkroom, a chin rest, etc.

It was not until David Katz's pioneer work on color constancy (1911) that fairly "normal" variations of situational circumstances (in this case of illumination), with a fairly representative array of (distance- and) illumination-cues were used, thus establishing the 1910's rather than the 1890's as the coming of

age of constancy research done in adquate manner. In this sense, modern experiments on thing constancy are deliberately "poorly controlled" with respect to cues, when viewed from the standpoint of the classical experimentalist.

More abstractly, the major difference between the two techniques may be formulated as follows. In the classical tradition, situational cues are gradually introduced, starting from one cue or factor which later—if not fully eliminated in an opposite alternative, reduction to zero—is enriched by the admittance of further factors in a process of increasing complexity that may be called "successive accumulation."[14] In contrast to this, the modern touch is given by a

TABLE 1

(Logarithmically recomputed data from original protocols of Holaday, 1933; averages from ten subjects, each in two distance constellations, 1: 2 and 1: 4 m.) *Illustration of the Technique of "Successive Omission" in the Analysis of the Effectiveness of Distance Cues in Perceptual Size Constancy.*—Binocular disparity seems fully dispensable (within the limits of statistical significance of the results) as long as other cues are present. Breakdown of the constancy mechanism does not occur unless the last remaining cue, relative motion, is removed by the prevention of head movements. Results from an additional subject, blind in one eye, show similar compensation.

	Average $c^{\log}$ binocular	monocular
Cubes with most favorable, "normal" conditions of the experiment	.88	.88
Cubes viewed through tubes restricting the visual field	.81	.79
Cubes barely discernible in natural twilight	.78	.74
Dim squares in darkroom, slight head movements	.75	.52
Dim squares in darkroom, head on chin rest	.67	.15

start from fully, or at least approximately, lifelike conditions, working downward in a process of decreasing complexity that may be called "successive omission" (in the sense of removal, or factual exclusion) of a cue or other condition. Instead of the term "successive" the adjective "fractional" may be more appropriate wherever accumulation, omission, or modification of several factors is effected at the same time by a certain change of conditions, as occurred in the study by Holaday as referred to below.

One of the great advantages of the omission technique is that all the vast, and to a large extent unknown, reservoir of so-called "minimal cues" can be preserved or kept in the situation much more readily than with the accumulation technique, and the effectiveness of some of the major factors can be studied against this natural background.

A special effort to demonstrate the omission technique for the constancies in general, and for the mechanism of depth perception in the establishment of size constancy in particular, was made by Holaday (1933). Some of his results, recomputed from the logarithms of the original data, have already been presented above. Others, especially relevant for the present discussion of the omission technique, are added in table 1. They show the great importance, but also the dispensability in the presence of even just rudimentary other cues, of binocular disparity.

[14] An outstanding example of this method was the "subtractive procedure" in the study of reaction time, established in the 1860's but later discarded for lack of consistency in the results (see Boring, 1929, p. 338; Woodworth, 1938, pp. 302–310).

It seems from these results, more clearly than may be possible from other types of sources, that the classical experimentalist used channeled mediation in his experiments primarily because he tacitly assumed actual mediation to be likewise one-track, or one-cue. This may have given him the illusion of proceeding in a representative manner. The intersubstitutability of cues in the over-all biological function of stabilized orientation shows the inadequacy of such a conception.

VI. Forced Representativeness of Stimuli in Experiments on Social Perception. Ecological vs. Populational Generality

The state just described—representativeness of "circumstances" (such as distance), systematic variation and co-variation of major experimental variables—is about the limit of representativeness achieved in perception psychology as long as it proceeded under the power of its own academic tradition, and to a considerable extent against resistances coming from this same source. To do better, research in the perception of physical objects had to borrow from an altogether different tradition born of the exigencies of a social psychology eager to do work of practical significance without having to wait for sanctioned academic methods to catch up with their problems.

1. Experimental Design, Type D. Example: Social Perception of Traits from Photographs

The special problem fitting into our line of argument is that of the social perception of such covert distal variables as intelligence and other personality characteristics in reduced-contact situations, involving especially the viewing of still photographs. It was developed in this country after the First World War, chiefly by applied psychologists who were interested primarily in personnel problems and counseling and were comparatively untroubled by the esoteric methodological scruples of systematic laboratory psychology.

The particular Experiment D (figs. 5 and 6) chosen as an illustration was conducted by the present writer (1945). It follows closely a pattern set by Cleeton and Knight (1924) but uses a more adequate sample of what may be called "social objects." Results as presented in the figures and discussed below are likewise quite similar in spite of the use of photographs instead of a live exposure of the social objects on a stage—which was the procedure of Cleeton and Knight.

The situation peculiar to social perception as compared with simple "physical"-object perception is that not only the subjects, but also the objects in the experimental setup, are persons. Although it may theoretically be possible to vary systematically a trait of the social objects, say intelligence (by arbitrarily choosing an equal number of persons for each of a discontinuous series of equally spaced precise points along an IQ scale of arbitrary range), such a procedure would indeed be strange from a research point of view, to say the least. It would become utterly absurd if not only variation but also co-variation would have to be made formalistically systematic. To make the analogy with classical psychophysical experiments complete, other personality features or external characteristics, such as will power, or the height of the forehead,

would either have to be made to vary—e.g., by artificial selection of the social objects—in perfect correlation with IQ (tie with neighboring variables, or channeled mediation, respectively), or else have to be held constant at a precise value. It is evident that such an experiment, if actually set up, would defeat its very purpose on account of the unnaturalness of the personality patterns (including the external characteristics) as compared with those in actual social reality. In retrospect, it throws light upon the glaring inadequacy of attempting a functionalistic psychology on the basis of the designs of classical experimentation.

Statistical sampling of objects—social objects, to be sure—along with the traditional sampling of "judges" (the "subjects" in the narrower sense of the word) is thus forced upon the researcher in perception for the first time. In a simplified form, this is schematically described in the fourth (last) horizontal section of figure 7. Traits other than intelligence, I, are labeled T. These, as well as one group of external features, B_1, are assumed to vary and co-vary representatively.

All correlations shown in figures 5 and 6 are based on a number of social objects of $N = 46$, referring to all the members of a subunit of the Army Specialized Training Program preparing for personnel work at the University of California. According to the best available evidence (including an intelligence test administered by the Department of Psychology at the University of California, see fig. 6) their IQ's are not as uniformly high as was sometimes claimed for such groups, but range from about 90 to about 140 in an approximately normal distribution.

The head portions of standardized frontal photographs of the 46 Army students were reproduced in random (alphabetical) order, with the ears as nearly touching each other as possible, in two rows with a combined total width of 18 inches, so that their features were fairly clearly discernible.

The actual Experiment D further contained—as may be done to great advantage—a number of systematic, in this case constant, features of the potentially "expressive" kind, such as remotely controlled identical posture and attitude ("standing at attention," which ordinarily is expected to control emotional expression as well), identical clothing, and, furthermore, identical conditions of photographing (distance, illumination). They are referred to in figure 7 by the symbol B_2 and by the bracket connecting with the sensory surface of the subjects.

In this way the analysis was thrown back upon the more intrinsic features of the social objects, "putting in parentheses" such previously settled trivia as the size constancies, and excluding some of the more extrinsic socio-economic variables as possible determiners of the snap judgments required of the subjects. In a general sense these systematic features sufficiently characterize the study as an experiment rather than a statistical survey; but even in this respect it retains the character of a modern rather than a classical experiment by the use of the technique of fractional omission rather than of fractional accumulation of the few systematic features, leaving the host of other variables representatively varied and co-varied.

The vast potentialities inherent in an experiment conducted in this fashion are only to a small degree exploited in the results shown in the figures pre-

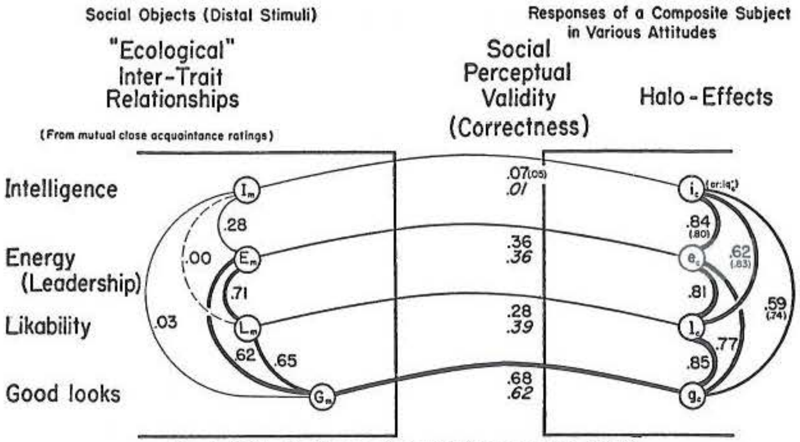


Fig. 5. (Data by Brunswik; preliminary report, 1945.)

Experiment D. Social Perception from Photographs: Functional Validity and Halo Effects.—All numerical results shown in this and the next figure indicate correlation coefficients based on an N of 46 (see text), with thickness of connecting lines (five steps) made to reflect the degree of correlation. The four variables listed on the left-hand side and designated by capital letters *I*, *E*, *L*, and *G* with the subscript *m* are based on averages of the ratings the soldier students—all of them closely acquainted with one another—were asked to give of each other; these "mutual" ratings or "opined" traits are the next best thing to the unattainable ideal of an "objective" appraisal of their personalities (see subsection 3). The symbol for goodlookingness is thereby moved somewhat to the right to indicate the semiexternal character of this trait. Lower-case letters at the right of the figure refer to four attitudes, corresponding to the four traits listed on the left, according to which the photos were to be viewed by a group of 25 regular students in a course in experimental psychology who were ascertained to be unacquainted with the Army students. As were the mutual close acquaintance ratings mentioned above, the intuitive estimates by these 25 subjects were made on a 7-point scale with special limitations minimizing the "generosity" factor and injecting an element of ranking into the procedure. Each of the resulting four response variables *i*, *e*, *l*, *g* which enter the correlation coefficients comprises a single "composite subject" (hence the subscript *c*) defined by the average ratings of all the 25 subjects for each of the 46 social objects separately, and in the attitude in question. Correlations with estimates of intelligence given in a "critical" attitude concentrating on IQ proper and thus labeled *iq'* appear next to those with *i* in small print and in parentheses. Correlations of both these types of estimates with two intelligence tests as specified in figure 6 are even less encouraging than those shown here; they are slightly on the negative side. (*Italicized* coefficients are based upon average ratings by a separate group of 30 subjects to whom the full-figure photographs rather than just the facial cutouts had been presented.) For coefficients based on individuals rather than the composite subject see subsection 2. Concerning reliability (agreement among mutual raters as well as among subjects) see subsection 3.

sented. Whereas, in the classical experiment, interference with actual conditions is such as to make evaluation beyond a certain minimum of information impossible, representative material can be evaluated under steadily changing new aspects, including an isolation of variables by such computational, after-the-fact techniques as partial or multiple correlation. Little of this has been carried out for the present example, but the statistical study on size constancy to be reported in § VII/3 has made inroads in that direction. The simple correlation coefficients presented reveal or reiterate fundamental facts of three kinds:

a) *Ecological relationships* of intelligence with some of the "mediating" external characteristics such as facial features and body build are surprisingly

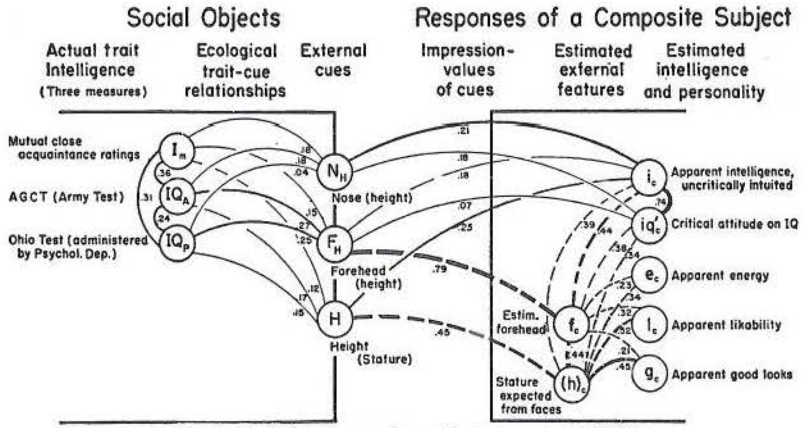


Fig. 6. (Data by Brunswik; preliminary report, 1945.)

Experiment D. Social Perception from Photographs: Problems of Mediating External Cues.—See legend to fig. 5. Again, the N for all coefficients is 46. Reference to two intelligence tests given to the Army students is included in addition to the mutual intelligence ratings of the Army students, and their (unexpectedly low) intercorrelations are given. Primary concern is with the relationship of the real trait intelligence (i.e., its three measures), as well as of various kinds of responses, to a few comparatively promising external physiognomic cues (see text). Some estimates of external physical characteristics by the subjects are included to show that halo effects extend even to the anticipation of actually not presented data such as the height of a person whose face alone is shown (emotional macro- and micropsia). The composite subject for this latter type of estimates consists of only 20 out of the total group of 25.

low. They do not surpass about .25 even for the physical features represented in figure 6, which were selected from a larger number investigated in our experiment as the ones most closely approaching significance (which would begin at about .3 for our sample of 46 social objects). Concerning the agreement of our coefficients with pertinent data in the literature on physiognomics see § VIII/1.

b) In consequence, the *functional validity* of physiognomic trait impressions based upon the reduced and static stimulus basis of a photograph (or, as in the study by Cleeton and Knight, even upon a static stage presence) is generally found to be rather low (fig. 5). In this respect intelligence is worst off, with correlations of practically zero. Personality features, such as energy and likability, are in our material just reaching significance, with coefficients in the neighborhood of .35. All this is quite in agreement with data reported in the literature (see below, § VII/1).

c) *Halo effects* as defined by Thorndike (1920), i.e., correlations between responses given in various attitudes, are surprisingly high, between about .6 and .85 (fig. 5).[15] They seem hardly justified even in a statistical sense in view of the lesser or completely vanishing corresponding ecological intertrait relationships among "neighboring" personality variables, at least for all combinations with intelligence where none of the coefficients surpasses .28. See also the

[15] The specific data from which the correlation coefficients representing the halo effects in figure 5 are derived were obtained with a carefully "balanced" order in which all 24 possible sequences of the four attitudes were used with at least one of the 25 subjects.

TABLE 2

TYPES OF MATERIAL INVOLVED IN VARIOUS USAGES OF THE CORRELATION COEFFICIENT RELEVANT TO THE PSYCHOLOGIST

(While in the rest of this paper quotation marks are often used to introduce a new term, they are here limited to designate accepted usages.)

Rows	General field of research	*Sample cases* and ("Individuals")	Universe ("Population") from which they are drawn	Original variable ("Trait," "Test"): Sample is exposed to *one*	Original variable ("Trait," "Test"): Character or region of values obtained	Original variable is correlated with (validated against)	Correlation coefficient describes	Conventional (or suggested) symbol
(1)	Differential psychology	*n Individuals* proper (Reacting organisms, Subjects; "Persons" as used by Stephenson)	Population proper	*Test situation* proper (Task, Object, Problem)	Trait proper: Responses *by* the subjects (or their total Scores; "Test" as used by Stephenson)	Repetition(s) of test (same or similar)	"Test reliability"	r_{11}
(2)						Other test(s)	"Test validity"	r_{12}
(3)	Functional problems (in the wider sense) and	*N Test situations* (Objects, Instances of varying Stimulation, Tasks, Problems; Persons in the case of special problems only, such as social perception)	Ecological habitat, or Culture	*Individual* (Subject, Judge) partaking of habitat or culture, and maintaining a certain attitude	Responses *to* the situations	Repetition with other individual(s)	Inter-individual observational reliability ("Agreement among judges")	$(r_{R_1R_2})$
(4)						Repeated trial(s) with same individual	Intra-individual observational reliability (Self-consistency of judge)	$(r_{RR'})$
(5)						Responses in other attitude(s)	Attitudinal validity ("Halo effect")	$(r_{R_aR_b})$
(6)						Stimulus variable(s) in situations concerned	Functional validity (General statistical correctness; Achievement)	(r_{SR})
(7)	Psychological ecology			Measurement or other "independent approach"	Stimulus variable in situations concerned	Measurement repeated at other time(s)	Ecological reliability (of a stimulus)	$(r_{SS'})$
(8)						Other stimulus variable(s) in situations concerned	Ecological validity (of a stimulus)	$(r_{S_1S_2})$

legend to figure 6. That halo effects may go so far as to include a person's name as a contributing variable to physiognomic impression is shown in such studies as that by Razran (1938).

Further data and discussion of the responses of individual judges will be given subsequently in this section. But first, the comprehensive scheme of the four model experiments repeatedly referred to above, figure 7, will be inserted here. Rather than taking too seriously the explicit verbal superstructure which is frequently overemphasized in the discussion of psychological "systems," the scheme represents an attempt to project the underlying ideology of the various schools onto the palpable level of experimental methodology.

2. CONVENTIONAL TEST RELIABILITY AND TEST VALIDITY OF SOCIAL PERCEPTION WITH REFERENCE TO THE JUDGES

The variables entering the correlation coefficients used in figures 5 and 6 to describe the results of the experiment in social perception are of a general type other than that used in the testing of individual differences among "subjects" when this term is limited to the individuals directly observed by the experimenter rather than to those designated as social objects in the present context.

In fact, use of average, composite ratings in those coefficients has all but obliterated inter-subject differences; the entire investigation may as well have been done with one subject only. In table 2 this type of approach is compared with the conventional testing approach—previously referred to in this paper (in § II/4 and in the discussion of model experiments up to and including C)—in terms suitable for further application to problems of physical perception as attempted in the subsequent sections of this paper.

In the testing of subject differences there is a sample of *n individuals* proper which is drawn from a population proper and tested for *one trait* which in turn is validated against another trait paired with the first in the individuals of the sample (row 2). This establishes what is commonly known as the "validity of a test."

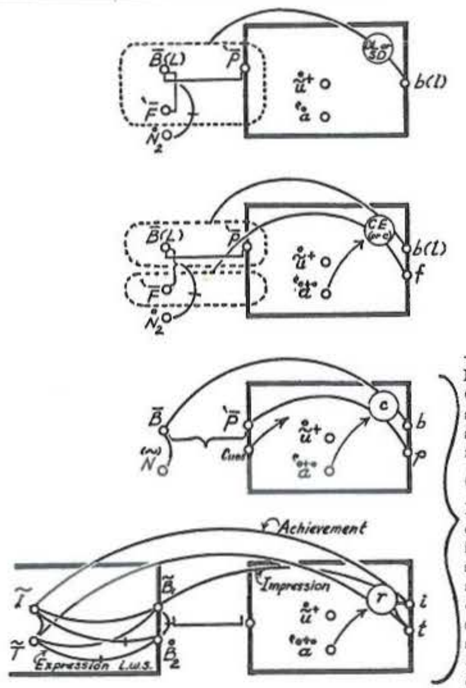
Any approach to test validity within our material on social perception requires restoration of interindividual variation ignored in the results presented in figures 5 and 6. One may, then, correlate the correctness of estimates of a particular social object (as the test situation) given in a certain attitude by the various judges with, say, their scholastic records. Instead of selecting one or another of the objects, one may use a measure of over-all correctness for the total of social objects as a composite score characterizing the social-perceptual achievement of each individual judge as the initial variable. Such a measure is given, say, by correlation coefficients computed in the manner of those shown in figure 5 for the 46 ratings by each of the 25 judges separately rather than for their 46 composite ratings. Such a set of 25 coefficients was actually computed for the attitudes toward intelligence and likability, as well as for their cross relationship constituting one type of halo effect in each judge. Among the correlations of these three sets of correlation coefficients with one another as well as with an index of scholarship derived from a series of course grades for each judge, we cite only the numerically greatest among them as an example. It is obtained by combining $r_{L_m l}$ as variable (1) and scholarship as (2) in an r_{12}, found to be $-.21$. Since the n for this latter r is only 25, the coefficient is far below significance, and so the slight negative validity of intuiting likability from photographs as a test for scholarship, indicated by the obtained coefficient, may not be taken for granted. The fact that intuitive perception even of a complex nature is generally not very closely related to intellectual performance is not new; witness the material quoted in the preceding sections.

If validation is against an "equivalent" form of the same test, or one "split half" of the test is correlated with the other, or when comparison is with a mere "retest" with the same

Experimental Design / Tradition	Affinity to School	Topical subdivision, and first vogue in adequate style	Handling of independent variables and formal character as psychophysical research	Turn of discovery (general trend of results), and role of organismic attitude (motivational set)
Systematic; European—Academic	Structuralism	Classical psychophysics (Sensory attributes, simple magnitude) Middle 19th Cent. (§ III)	Alleged isolation of one variable; actually inconclusive within cluster, tied by channeling of mediation, etc. Proximo-distally neutral, pseudo-unidimensional psychophysics	Remarkably fine discrimination (threshold, difference limen DL) under these favorable conditions Attitude not an issue (except for general attentiveness, number of categories of judgment, and other nondirective aspects)
Systematic; European—Academic	Gestalt psychology	Dynamics of form perception (Interaction within field) 1890's (§ IV)	Systematic isolation of figural variable from one of neighboring variables by interlocked co-variation Proximo-distally neutral multidimensional psychophysics	Remarkable contributingness of neighboring and organismic factors: large illusions (constant error, CE) Importance and flexibility of variable-directed attitude in field organization
Representative; American—Social and Applied	Functional approach	Thing constancies (Intuitive orientation among physical objects) 1910's (§ V)	Systematic isolation of distal from major mediating variable; other cues representative ('poor control') (Overt-) distal vs. proximal multidimensional psychophysics	Remarkably stabilized relations to physical objects regardless of sensory mediation (large constancy ratio, c) Comparative fixation of attitude on distal focusing
Representative; American—Social and Applied	Functional approach	Social perception (Intuitive orientation in social environment) 1920's (§ VI)	Statistical isolation within network of representatively varied and co-varied factors; farthest environmental reach (Covert-) distal multidimensional psychophysics	Fair achievement (r) for emotions, less for permanent traits, in static, reduced contact Moderate attitudinal shift span, strong contributingness of neighboring variables (halo)

Fig. 7 (on this and the opposite page). *Comprehensive Schema of Four Types of Experimental Design in Perception.*—The drawings represent the organism as a physical system in its physical environment in a way similar to the present writer's graphic treatment of the regional texture of emphasis, or "conceptual focus," of psychological schools (Brunswik, 1939a, further specified 1952). The drawings are therefore not to be confused with Lewin's (1935, 1943) representations of the "life space" which is the internal dynamic field reflecting the organism-environment situation and therefore to be located in its entirety within our layer *O*, possibly as an aspect of the temporary pattern of *a*. Only the barest minimum of the variables involved, necessary to bring out the basic points of experimental design, is shown in each schema. The school designations along the left-hand margin are to be understood in broadest terms of research policies and practices as operationally reflected in research methodology. Especially the term "functional approach" is meant to refer primarily to the general current trend toward an emphasis on the biological adjustment value of behavior (in our case, on "orientation" in the environment and its over-all correctness) rather than to the by now historical Chicago "functionalist" school of the turn of the century. Though first formulated in this latter movement, such an emphasis was there but incompletely understood in its concrete research implications. (For references see the Introduction). The four diagrams, encompassing the full span of the history of experimentation in perception, reflect the increasingly broad reach of the arch connecting the variables of reference of experimental research, as well as the increasing complexity of the surrounding and intervening patterns of investigation.

Example chosen, variables confronted, and specific trend of results	Regions in perceptual experiments (§ I): Ecological variables — Distal stimuli: covert S_C, overt S_D; Proximal stimuli S_P; (Central-) Organismic variables O; Response variables R
Experiment A: Galton bar Most closely akin to measurement of length as a physical stimulus variable Dominance undecided	
Experiment B: Müller-Lyer illusion *N*-crucial alternative: Length (*L*) vs. fringing area (*F*) *F* slightly dominant	
Experiment C: Size constancy *M*-crucial alternative: Bodily size (*B*) vs. visual angle (*P*) *B* decidedly dominant	
Experiment D: Traits from photos Many *N*- and *M*-crucial alternatives Poor attainment of *I*, some of character traits	



Further development: Convergence of academic content and social and applied method (§ VII)

Example: 'Survey' of size constancy in the manner of social perception—a methodological demonstration of stimulus-response statistics. Confirms distal focusing of perception

Signs for methodological concepts used in this figure (§ II)

	Variation		*Co-variation*	
Systematic design (straight lines, angles, dots)	Systematic	'bar'	Artificially tied (unison)	'bracket'
	Few constant values	'double dot'	Artificially interlocked (semi-representative)	'brace'
	Constant (or zero)	'dot'	Artificially untied	'bisected arc'
Representative design (curved lines)	Representative	'tilde'	Representative:	
	Quasi-constant (mean of sample)		Intra-ecological	('hanging') 'arc'
			Functional	('standing') 'arc'

Modifiers: Added experimental unit (between symbols) / Added evaluation (after a symbol) — 'add'

Remote control (through antecedent condition): Rigid, univocal — 'mirror prime'; Statistical (probable) — 'mirror apostrophe'

Condition limited (with graphic symbol) — 'parentheses' group of individuals, the resulting coefficient is said to represent the "reliability of a test" (row 1). The only instance in which our data on social perception come close to furnishing material on equivalent attitudes in the same group of subjects viewing the photographs is in the case of *i* as compared with *iq'*. Proceeding on an individual basis in analogy to test validity, and using $r_{I_m i}$ as (1) and $r_{I_m iq'}$ as (I), r_{1I} was found to be .54. This is significant even with our small n of 25, although by no means very satisfactory by customary standards.

3. FUNCTIONAL VALIDITY, HALO EFFECT, AND OBSERVATIONAL RELIABILITY

The type of relationships studied in subsection 1 of the present section takes up the rest of table 2. They have in common that they are in terms of an ecological sample of *N environmental situations* or stimulus objects (in our case, social objects), each constituting a separate test, task, or problem which is drawn from a universe or supply of such situations. In our special case, this universe still happens to be a population proper which, however, is not the same as that of the judges. Generally, such supplies are constituted as part of the ecological universe of *one individual*, species, or culture. In our case, this is a composite judge, or any one of the individual judges; in our statistical survey of size constancy to be discussed in § VII/3, it will be actually only one judge.

A first type of variable to start from is given by the responses, made by the one individual just mentioned maintaining a certain attitude, *to each item* in the sample of situations or objects (rows 3 to 6). Following the analogy consistently, this is a trait variable of the objects rather than of the subjects (for further explanation see below).

While in test reliability and test validity it is likewise responses which furnish the starting point for correlation, these responses are *by each subject* in a sample, to a certain fixed stimulus situation, and thus constituting a trait variable of the subjects. (Each single response—say, an estimate of +3 by subject i given to object k—then pertains to that one subject and one object only and is thus a trait *value* of both of them at the same time.)

We may start our discussion with what is perhaps the most familiar in this group of usages of the correlation coefficient: halo effects (row 5). One type of response is here validated against a response given in another attitude. The coefficient thus indicates what may be labeled "attitudinal validity" (to be subsumed under functional validities in the wider sense of the term). The correlation coefficients along the right-hand margin of figure 5 all indicate this type of relationship.

In case the response variable is validated not against another response but against an antecedent stimulus variable (distal or proximal) established within the same sample of situations, the term "functional (in the sense of observational) validity" or "perceptual achievement"—or also "ecological validity of a response," cf. subsection 4—may be used (row 6). Thereby "achievement" has previously been defined as a "generic term" designating "the relationships better than chance existing between, and due to, an organism and variables in its physical environment" (Brunswik, 1943), in short, over-all statistical "correctness" rather than single hits or misses of judgment.

Of special importance is the case in which validation is against the variable "intended" in perception, or, as we may also say, the "homonymous" variable,

i.e., the one bearing the same verbal or letter designation as the response (e.g., *I* vs. *i*). Examples are found in the middle portion of figure 5. Validation may of course also be against a "heteronymous" stimulus variable, such as against external physical features as in the middle portion of figure 6.

Whereas representative perceptual achievements in social trait perception are generally low, they are higher for emotions (see § VII/2) and especially for physical size constancy (see § VII/3).

Moving from halo effect upward in table 2, we come to the discussion of a type of inter-response correlations which are of special importance in the establishment of the concept of "objectivity" of observation. There are two subvarieties, to be labeled "intra-individual observational (or functional) reliability" (row 4) and "inter-individual observational reliability"—also to be called "ecological reliability of a response," cf. subsection 4—(row 3). The former refers to the self-consistency of a judge, or to what may be called maintenance of attitude and/or response from one time element to another, in short, to the agreement of a judge with himself in repeatedly responding to a certain sample of situations. It is the kind of variability that also underlies the theory of errors of measurement or of psychophysical judgment referred to in § III/2. The latter subvariety of observational reliability does the same for agreement among two different judges; it is thus the only instance other than test reliability and test validity in which more than one subject is brought into the picture, and even so the scheme does not provide for more than a pair of subjects at a time rather than an entire sample of them. And, what is more, the procedure is not in itself diagnostically relevant with respect to the subjects doing the rating.

The closest approximation to a measure of self-consistency of our composite judge available in our material is given by correlating i_c with iq'_c. As reported in figure 6, this coefficient is as high as .74, but still less than some of the halo effects shown in figure 5. Separate coefficients of this kind for each of the 25 subjects have not been computed.

Inter-individual agreement is even more important in the literature referring to ranking and rating than is intra-individual agreement. So far as our own data are concerned, it may suffice to give the values for the trait intelligence. Using an adaptation of the short-cut method described by Woodworth (1938, pp. 373 f.), the average agreement among all possible intercombinations ij of the 25 subjects viewing the photographs in attitude *i* has been found to be .26. (To be sure, we are dealing here with the judges as a sample; but the value given is a composite r—to be distinguished from a single r derived from composite judgments—which may be reduced to a number of single r's combining the individual judgments given by two judges.)

Considering for a moment the soldier students as raters of each other rather than as social objects, a similar average coefficient of inter-rater agreement may also be computed for I_m. With 41 of the 46 social objects participating, the average r was found to be .33. This is hardly more than the value obtained for the subjects, and certainly not enough to establish "objective measurement"; however, it probably still remains the best approximation to objectivity available to us without going too much out of our way. Tests are of course valuable, but interrelationships among the instruments as shown along the left-hand margin of figure 6 are as disheartening as is inter-rater agreement. The low value of the latter in itself imposes a limitation upon inter-instrument agreement for two of the three combinations, those with I_m. It may further be argued that the concept of objectivity should not be based on inter- or intrasubjective agreement exclusively, but also on what is sometimes called "independence" of the approach (see Brunswik, 1952). Mutual close acquaintance ratings, regardless of their observational reliability, are certainly independent of the subjects' ratings both as far as the personal identity of the raters as well as, what is more, the overwhelming portion of the stimulus basis on which the ratings are made is concerned. In this latter respect, the stimulus basis available to the social objects is very considerably greater than that of the subjects.

Intuitive ratings of personality variables based on long-range or otherwise close acquainteanceship are becoming increasingly important in the recent dynamic psychology of personality, as shown in such recent publications as Frenkel-Brunswik (1942). Although the number of raters may for obvious reasons be quite small, the problem of inter-rater agreement can nonetheless be approached even from a differential angle in addition to problems of intra-rater consistency.

All four uses of the correlation coefficient discussed in this subsection may be said to deal with functional problems in the wider sense of the word, i.e., with various aspects of organisms functioning in the face of a variety of external situations. For all of them it may be said that in a certain way individuals and test situations have shifted places (see table 2). Yet this type of exchange has to be sharply distinguished from the "inverted" correlation technique recently introduced by Stephenson ("Correlating Persons Instead of Tests," 1935; see also 1936). By accumulating tests for paired intercorrelation in such a way that their number, m, becomes comparable to the size of the sample of subjects, n, that is, by giving many tests to many individuals, factor analysts arrive at matrices constituted by columns representing persons and rows representing attributes (tests, traits, measures, features) of these persons. This holds true regardless of further evaluative manipulation of the data represented in the matrix. In inverted correlation the persons are used as variables and the tests are the population (they are usually not really a sample in the usual sense, see § IX/1), with "types" as final result—instead of vice versa as in the customary type of factor techniques used by Spearman, Thurstone, Kelley, or Hotelling. Thus, inverted correlation is a much less radical departure from the customary ways of psychological statistics than the application of correlation and sampling aspects to functional stimulus-response problems as suggested above. The latter is object-centered rather than subject-centered in the sense that there are no "persons" in the plural at all (except in the role of distal stimuli; see especially also § VII/3), but only a sample of external situations responded to, on principle, by one subject.

There is no difference in the type of *material* used by Stephenson and the other factor analysts. The only difference is that correlations are calculated between columns, rather than between rows, of a common "persons × their features" matrix. If our own case were extended, as it should be (see below), to integrate more than two (say, M; see below, subsection 4) features of a sample of situations at any one time, the matrix would be one of "situations × their features" (i.e., using, as we have already done, N for the size of an ecological sample, an N × M matrix) rather than one of "persons × their features" (i.e., an n × m matrix). In the end, one may thus arrive at a factoring of perception in terms of situational dimensions. This would be a counterpart to the already well-developed customary type of factoring in terms of personal abilities as recently also extended to perception by Thurstone (1944, see above, §§ IV/2 and V/1; see also Thouless, 1938). But it would be diametrically opposite to the latter with respect to type of content since there would be not merely an exchange of rows for columns, as by Stephenson, but a substitution of an altogether different set of variables for those which enter into the matrices dealing with differential psychology, be it that these matrices are used for the customary kinds of factor analysis or for the Stephenson technique.

4. INTRA-ECOLOGICAL CORRELATIONS

The difference between differential and functional approach disappears only when we proceed to further abstraction. Subjects and physical or social objects, as well as situations in general, are but aggregates or sets of condi-

tions; the situations may be considered as momentary or as encompassing segments of time by including causes or effects of these conditions. Universes of such aggregates may be defined by fixed or quasi-fixed criteria which may or may not include reference to a responding subject. Variable features or traits of samples drawn from these universes may be recruited from among the conditions extant in these aggregates that are permitted to vary, or they may be newly provoked by exposing the aggregate to a probing instrument or test situation which thus becomes a fixed addition to the total set.

If this instrument is a human subject, his responses, or, generally, effects to which he has contributed in the context given, become features of the sample of situations involved. In this sense I_m, the two tests of IQ listed in figure 6, as well as *i* and *iq'*, are but five features of the social objects and thus, loosely speaking, five tests of the environment. For the last two of them, mediation is limited to special effects of the social objects upon photographic plates, and it is the probing subject which constitutes the quasi-constant test condition relative to the social objects. The number of impersonal or personal tests thus given to a sample of environmental situations may of course be extended much farther than has been done above; it may be designated by the letter M, in analogy to the letter m which is the customary designation for the number of tests given to a sample of human beings.

In conventional testing, on the other hand, a sample of subjects is used as aggregates of conditions, and probing is by a tester who has become impersonal by being the mere administrator of a more or less rigidly fixed test instrument approaching "measurement" as close as possible by possessing the highest obtainable observational reliability (in the sense defined in rows 3 and 4).

A third group of uses of the correlation coefficient in addition to those in testing and in functional analyses emerges by depersonalizing, in the manner of measurement, both the varying situations and the probing instruments. A remnant of subject reference may be maintained by defining the universe from which the situations are drawn in terms of an organism the environment of which they constitute, thus rendering them "ecological" in the sense defined in § I/1. By relating one feature of the environment proper to another, we obtain coefficients designating what may be called "ecological validity of a stimulus" (row 8).

An example of a purely intra-ecological analysis of physical-environmental variables, relating potential distal to potential proximal stimuli, is given in § VIII/2, with reference to a current investigation of the objective validity of depth criteria. Although this analysis is novel in content, as is that of the functional validity of responses with respect to physical-environmental variables, the objective, intra-ecological validity of physiognomic features, relevant for social perception and discussed in § VIII/1, has repeatedly been approached statistically within the last few decades.

Correlating environmental features with themselves by comparing measures at two different time elements with one another leads to what may be called "ecological reliability of a stimulus" (row 7). As long as we uphold the ideal of perfect observational reliability as defined above, the application of different measuring instruments to a common hypothetical variable such as "intelligence" may not properly fall under this classification. In the ideal sense, ecological reliability should refer only to the objective constancy of features which are univocally measurable at any moment. In this sense such features as the physical size of rigid bodies (as long as these do not grow or shrink appreciably) or the "albedo" of their surfaces (except as in, say, the chameleon or a blushing adolescent) are ecologically highly reliable and thus a challenge to organisms in need of Archimedic points (or Archimedic variables, if one may say so) to support orientation in the environment; whereas retinal sizes, or the wind, are examples of changing, ecologically unreliable variables with usually lesser orientation value.

5. SAMPLING-OF-SUBJECTS RELIABILITY AND SAMPLING-OF-OBJECTS APPLICABILITY. ECOLOGICAL VS. POPULATIONAL SIGNIFICANCE OF RESULTS

Correlation coefficients are summary statements of results which are descriptive in the sense that they do not in themselves go beyond the actually observed evidence. Yet by being summaries they contain, at least implicitly, the claim to generality. The degree to which this claim is justified is usually indicated by an added reference to the standard error (SE), so that we may anticipate the reasonable limits within which coefficients from other samples may be expected to lie.

Since, in the preceding pages, sampling was not only considered for subjects but also for objects, it is clear that the statistical significance of a result may be investigated in both these directions. In particular, we may ask ourselves not only the customary question: Which social perceptual achievement (or halo effect) may we reasonably anticipate when presenting the same set of photographs to a new group of judges?—a question which refers to what may be called "populational" generality of the result; but we may also ask the additional question: Which achievement may the same judges (in the same attitude and assuming that there is no practice effect) be expected to display when judging the photographs of a new group of social objects? With respect to this latter query it was suggested (Brunswik, 1944) that we should speak of the question of the "ecological" or "situational" generality of the results. Ecological generality or significance is actually based on a kind of sampling-of-objects reliability, and may be derived from what may be designated as SE_{ecol}, in contradistinction to the conventional standard error which may be specified as SE_{pop}. Ecological generality may also be characterized as "sampling applicability," with special consideration of applied psychologists who justly complain about the nonapplicability of the results of most classical experimentation to their problems.

Let us give only one concrete example of the two types of significance. The halo effect between *i* and *l* for our composite subject was found to be .62 (see fig. 5). With an ecological N of 46 and therefore with an SE_{ecol} of .09 for this coefficient, the existence of a halo effect appears ecologically significant, with the composite subject held constant. On the other hand, the mean of the individual correlations between *i* and *l* for the populational n of 25 subjects who make up the composite subject is .57 (somewhat lower than the group correlation; see § VII/1), and the SE_{pop} of this mean was found to be .04. Populational significance, with the sample of objects held constant, is therefore likewise ascertained although at a different level than is ecological significance; this is achieved by treating the correlation coefficients, characterizing the halo effect displayed by the various subjects, as individual scores. By further adjustment of the technical concepts involved it may also be possible to apply the two types of SE and the subsequent criteria of significance to the same measure, rather than once to a coefficient based on mean ratings and once to a mean coefficient based on individual ratings as was the case here.[15a]

Ecological generality of experimental or statistical results may thus be established along with populational generality. In fact, proper sampling of situations and problems may in the end be more important than proper sampling of subjects, considering the fact that individuals are probably on the whole much more alike than are situations among one another.

In a broader sense of the term, all considerations of validity imply variation of the ecological scene and thus have to do with ecological generality. Both sampling-of-objects problems as well as those of validity may therefore be included under the term "applicability" in a broader sense, just as the term "reliability" customarily extends to both test reliability as well as sampling-of-subjects reliability, aside from the further uses listed in table 2. But whereas test validity only permits prediction of the probable position of a testee relative to other testees on a new test, ecological sampling applicability permits prediction of his probable position on a new test as described in terms of an absolute score.

VII. Convergence of Experiment and Statistics in the Methodology of a Probabilistic Functionalism

1. TRADITIONAL DOUBLE STANDARDS FOR THE SAMPLING OF SUBJECTS AND OF STIMULUS OBJECTS

The equality of the status of subjects and objects with respect to representative sampling just postulated may seem trivial. It must appear reasonable after having been explicitly suggested. Yet, since science is not only a matter of form and explicit logic, but also a matter of specific content and of implicit habit, historical progress in this direction has been slow. The very use of the terms "individual" or "population" in the manner indicated in the head portion of table 2 seems to create a frame of mind concretizing the channels of investigation and discussion in the direction of problems of individual differences among subjects.

The present writer has in himself experienced the required shift of emphasis as very slow going and hard to maintain, especially so far as consistent concrete application is concerned. The difficulties he encountered in explaining his point, after he had succeeded in at least establishing a bridgehead for himself in abstract terms and in a few concrete examples, have given him the impression of resistances approaching in intensity those encountered in the opening up of emotionally highly loaded topics, such as those dealt with in psychoanalysis. Indeed, the deliberate abandonment of "systematic" policies, especially where such policies are technically feasible, in favor of statistical practices which by comparison seem slovenly, must on the surface seem to violate one of the most fundamental taboos developed in the traditional ideology of the "exact" sciences.

[15a] As has been suggested by Professor Jarrett while reading the proof, the two types of measures might be merged into a single test of significance. Although there can be no doubt of the interwovenness of the two aspects, it should, however, be emphasized that the ecological components—coupled with scrutiny of the adequacy of representation, in the research design, of ecological dimensions along with the populational ones—must remain one of the primary concerns of the investigator.

Even those among the researchers who are relatively unburdened by established intramural routines and to whom we are indebted for first tackling the problems of social perception on an inductive basis have not always felt the need, in their research practices, of paying to the sampling of social objects an attention equal to the attention they would, as a matter of course, pay to the sampling of judges. In the typical case, they have a pitifully small ecological N to go with a populational n of adequate size. In keeping with our Experiment D discussed above, we shall refer primarily to studies endeavoring to expose the low validity of snap judgments of intelligence.

Thus Pintner's study (1918)—one of the earliest and most naïvely executed in the field—uses only 12 social objects, ranging unrepresentatively from extreme brightness to feeblemindedness and from 4 to 16 years of age, with conditions of photography far from standard; all these are features hardly in keeping with his adequate array of 63 judges. Gaskill, Fenton, and Porter (1927) limit age to 11–12 years but have likewise only 12 social objects with IQ's ranging from 18 to 171, as contrasted with judges numbering 274. Estes (as late as 1938), in an otherwise broadly conceived, sophisticated study of personality patterns, uses films of only 15 male students, viewed by a total of 323 judges. Viteles and Smith (1932) present photographs of 10 professional men, 5 of them successful and 5 of them unsuccessful, to large groups of personnel experts as well as of untrained students attempting to estimate successfulness. And, more drastically than in any of the other cases, Laird (1925) refers to a newspaper report according to which a "crowd" tried (unsuccessfully) to tell a railroad president from a feeble-minded criminal seated on a stage—2 social objects in all—as proof of the general futility of judging intelligence and personality from external appearance. In reports on his own additional efforts he refers to one or two samples each of 5 and 10 photographs, the latter having been given to some 400 judges.

It is clear that sampling-of-object-reliability, i.e., sampling-applicability in the above-defined sense, of correlation coefficients or other measures of correspondence based on such small samples of objects is extremely limited. On the other hand, the largeness of the samples of judges is at least in part wasted by the averaging of their ratings. This holds in spite of the fact that, as pointed out by Hollingworth (1916, pp. 41 ff.; 1922, pp. 34–43), higher coefficients are obtained from group averages establishing a single composite subject than by averaging the correlations from the same judges treated as individuals. Although Hollingworth reports that the average of the coefficients of correctness of 10 subjects judging intelligence from photographs of 20 persons is only .19—their range being from −.27 to +.51—he finds that two "group judges" combining the verdicts of 25 men and of 25 women before correlations are computed average as high as .51, a figure otherwise unheard of in the study of social perception of intelligence from photographs. However, as reported by Hollingworth himself, the validities of judgments of intelligence found by Pintner (see above) and by Anderson (see below) are raised to a mere .16 and .27, respectively, by first combining the judgments. Furthermore, our own coefficients reported in figure 5—which are based on a sample of 46 photographs as contrasted with Hollingworth's inadequate number of 20 and are likewise computed from group averages—yield coefficients of only between zero and .1. This latter finding is quite in line with Cleeton and Knight's (see below) average correlation coefficient of .02 for estimates of "intellectual capacity" (paralleled, to be sure, by .32 for estimates of "soundness of judgment").

A major exception to the customary neglect of proper social-object sampling is a study by Anderson (1921) in which 12 judges were to estimate from printed photos the intelligence of 69 employees of a business firm. Even Cleeton and Knight (1924) have only three separate groups of 10 subjects each (two of women and one of men), to their 70 judges rating them on eight traits ranging from intelligence to impulsiveness.

Countless other examples from the field of social psychology, and especially from other fields, could be given for the inequitable neglect of stimulus sampling as contrasted to subject (responder) sampling.

2. CORRELATION AND FACTORIAL DESIGN IN SYSTEMATIC STIMULUS-RESPONSE ANALYSIS

So far as academic psychology proper is concerned, two steps may be distinguished in the gradual transition from systematic multidimensional psychophysics to a fully representative stimulus-response psychology. They will be discussed in this and the next subsection.

Among the concepts borrowed from traditional statistics to describe stimulus-response relationships the correlation coefficient is the most outstanding. Since experiments on the social perception of traits are, in certain respects at least, almost inherently representative, the correlation coefficient has from the very beginning been commonly used in describing their results. Yet, characteristically, in the somewhat more academic tradition of the social perception of emotions by means of their "expressions" (see fig. 7) the establishment of the correlation coefficient as a measure of correctness of judgment is relatively new.

One lone instance of such a coefficient is given in Woodworth (1938, pp. 250 f.). It refers to a rather artificial situation of guessing of emotion from photographed poses of an actor, based on a study by Feleky. It is interesting to note that for such temporary states of mind social perceptual validity may reach the extremely high value of .92, in sharp contrast with the very modest achievements with respect to the permanent traits discussed above; in part, the high value obtained may well be due to a factor of conventionality in the poses, and to the admittedly crude arrangement of the various kinds of emotion under investigation along a single dimension.

In the field of the perceptual constancies proper, the correlation coefficient was first used as a measure of correctness by the present writer (1940). It must be realized, however, that such an approach was statistical no more than in a superficial, descriptive sense. The more representative survey (Brunswik, 1944) referred to in subsection 3 was thus undertaken.

Another, more general device which has become prominent in establishing statistical thought in experimentation proper is "factorial design" in combination with analysis of variance.[16] It has given new impetus to the designing of experiments in psychology which are multidimensional and thus capable of probing interaction within a dynamic context; over and above this important step toward functionalism, however, they may well remain, and have so far largely remained, systematic, formalistic (in the sense of these terms as used in the present paper), at least as far as the variation and co-variation of the more easily controlled among the ecological variables is concerned.

In animal psychology a first experimental approach of this kind has been undertaken with string pulling in rats (Crutchfield, 1938, 1939; Crutchfield and Tolman, 1940). Five variables—all ecological—were used: force of pull (in training as well as in the crucial trial), length of string, training period, and time since feeding (the latter as a remote control on the motivational variable "hunger"). All possible combinations of three "levels of strength" chosen by the experimenter for each variable yielded 243 different conditions, each of them represented by a single rat. (This singleness is of course not to be confused with the single subject used in the present writer's statistical survey of size constancy, 1944; see below).

[16] Fisher (1925, 1935); summarized for psychologists by Garrett and Zubin (1943), with comments by Grant (1944). For translation of the terminology from the original field of agriculture to psychology see Baxter (1941). See also Lindquist (1940).

The systematic character of the general design of this experiment may best be visualized by a five-dimensional checkerboard, of the pattern 3 × 3 × 3 × 3 × 3. Since the arrangment leads to an even distribution of values throughout all the cells of the scheme, intercorrelations among the five independent variables have arbitrarily been made to disappear so that the variables are artificially untied from one another in the sense defined in § II/3. One may compare this with the design presented in the right-hand part of figure 4, in which the somewhat different technique of interlocked co-variation has led to an almost complete disappearance of ecological correlation, likewise without resorting to the holding constant of one or the other of the variables involved. Although the question of the representativeness of the three levels of strength of each factor was in the present example handled with no more than the customary casual implicitness, the lesser degree of artificiality inherent in factorial design as such is explicitly emphasized by the authors mentioned (1940). This is of course fully justified since multidimensionality as such constitutes an important step in the direction of representativeness.

An informally evaluated 3 × 3 × 3 × 7 design involving systematic variation of the (socio-)ecological variables forehead, eye distance, height of mouth, and height and position of nose (with a preliminary attempt to ascertain some degree of representativeness in the drawings) was used in an effort to study the physiognomic impressions elicited by schematized drawings of faces (Brunswik and Reiter, 1937, advance report in Brunswik, 1934; a reproduction of some of the drawings from the advance report may be found in Allport, 1937, p. 483). The use of three or more levels of strength for each variable permits ascertainment of curvilinear trends and is therefore indicated wherever such trends are suspected. The over-all averages of the responses of 10 subjects, each of whom did about twenty hours of a complex procedure of ranking the 189 drawings in seven physiognomic attitudes, show a general curvilinear tendency in the graphs, pointing toward a preference for the "golden mean" (middle value) for all features High forehead is likewise favorable throughout, whereas high mouth, short nose, and in certain respects also eyes wide apart are ambivalent in the sense of giving the impression of happiness and youth, but also of lack of energy and intelligence.

An attempt to check upon the results of Brunswik and Reiter was made by Samuels (1939), using 247 (instead of the original 10) subjects. They were, in turn, given to view only those ten of the schematic faces which had been found most extreme in expressiveness by the original investigators. Within these ecological limitations, the populational generality of the results of the original study by Brunswik and Reiter was clearly substantiated.

Samuels' approach, in a further experiment reported in the same article, to what is here called ecological applicability, however, has all the earmarks of the widespread unawareness, referred to above, with respect to the formal similarity of this problem to that of subject sampling. For each of the ten schematic faces considered, a "real" representative—only one—was reproduced from an album of photographs of Harvard alumni after the proportions of the variables involved had been established as identical with those of the schema to be duplicated in this more lifelike fashion. (Thereby one of the two aspects of the fourth variable was left out of consideration. It is the position of the upper end of the nose which varies considerably in appearance but hardly at all in terms of the accepted anthropometric measure, i.e., the position of the "nasion"—the juncture of the nasal with the frontal bones—relative to the eyes.) An inspection of reproductions, given in the article, of two of the photographs thus chosen reveals the extremely incidental nature of all the remaining features, such as color of hair, lines in the face, and even clothing. For a real check, either a large sample of faces for each of the ten drawings, averaging out incidentals, or else one or a few faces for each of the 189 drawings, would have been required.

Another experiment by Brunswik (1939c) with schematized full figures has been modified in a study by Wallace (1941) employing a partly systematic technique of enlargement, reduction, and distortion of photographs of actual persons so as to increase ecological applicability of the results. A guarded statement—caution being dictated by the small number of original social objects rather than on general methodological grounds—concerning the appreciable but relatively small contribution of height or weight to apparent intelligence and personality was thus made possible in the face of objections raised in § IX/3 against generalizations, in fully systematic experiments, regarding the relative strength of ecological factors.

In a vein complementary to that of the present paper, Shen (1942) deplores the "assignment of individual differences to the role of control rather than independent variables" in psychological experimentation. He discusses a factorial design combining 2 methods of practicing handwriting (an ecological dimension) with 5 age levels (a populational dimension), with the primary purpose of studying "interaction" between age and method of practice suggested by a previously established significant regression of the difference between the two methods upon the age of the subjects. Again, this experiment remains systematic in the sense that the two methods—free practice and training—are arbitrarily selected from the reservoir of possible training methods (see § IX/1). The remarkable reversal from an original emphasis on an ecological to a derived emphasis on an individual-differences content as proposed by Shen once more illustrates the primarily systematic, functional-experimental and only secondarily differential-statistical affinities of the method of factorial design in psychology.

3. DELIBERATE ECOLOGICAL (STIMULUS) REPRESENTATIVENESS IN A STATISTICAL SURVEY OF PERCEPTUAL SIZE CONSTANCY

From the standpoint of the classical experimentalist interested in the establishment of universally valid general "laws" rather than just imperfect correlations, the statistical approach must seem fundamentally inferior to the experimental. Statistics may seem acceptable as a necessary evil, or a temporary expedient, wherever one thinks he cannot do better, to be replaced by a systematic policy wherever and as soon as possible. From such a point of view, the experiment in social perception discussed in the preceding chapter must appear doubly deplorable since both stimuli and responding organisms had to be handled statistically.

From the utilitarian-functionalistic point of view espoused by the present writer, the statistical touch is an asset throughout. What remains to be done is to see as a virtue what was considered the calamity of imperfect control. This can be done by explicitly sanctioning, and especially by actually putting to work, a fusion of the fresh start in methodology effected by Anglo-American statistical applied and social psychology and reflected in the studies on social perception, with the content specific to the academic tradition of European experimental psychology (including its continuation in America), reflected in its most complex and up-to-date aspects by such a problem as that of the perceptual constancies.

An experimental-statistical hybrid study of this kind was undertaken by the present writer (1944) primarily as a methodological demonstration but also with an eye to ascertaining the ecological generality of certain basic principles stated previously in the field of the perceptual constancies or of perception in general. It is a study of systematically controllable overt distal physical perception, done deliberately after the fashion of not systematically controlled covert distal social perception.

Figure 8 presents a scattergram of the ecological variation and co-variation of the two independent variables primarily scrutinized, B and P. They are the same as those "confronted" with each other in Experiment C on size constancy. This time, however, they were not systematically varied in accordance with a preconceived plan of an experimenter, but randomly sampled from the normal environment of a university student, stopped in her daily routine by an

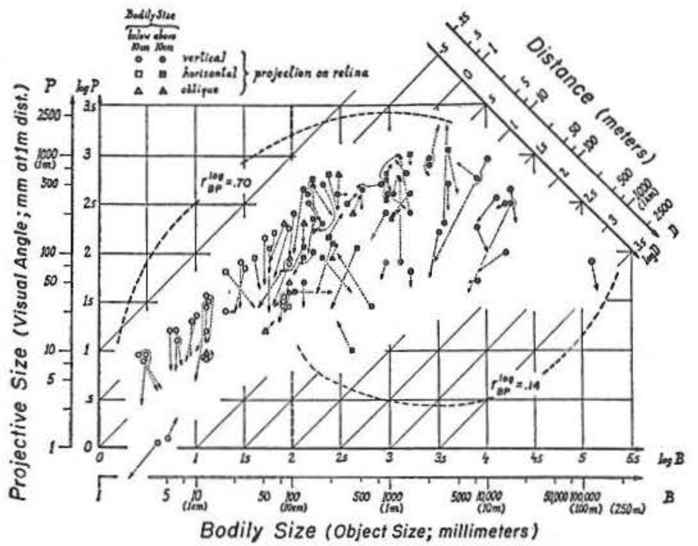


Fig. 8. (From Brunswik, 1944, fig. 2, with coefficients added from table 5.)

Scattergram of Ecological Interrelations of Object Size, Retinal Size, and Distance.—The large solid and outline circles, squares, and triangles describe the stimulus situations in terms of B, P, and D. The arrows issuing from these symbols connect with points based on the corresponding subjective estimates b and p whereby the homonymous axes are used as coördinates, thus proceeding to a representation of the responses of our subject, and indicating direction and magnitude of errors made in two attitudes with respect to two stimulus variables. (Regarding problems of consistency with the third response, d, see the original monograph.)

"experimenter" who was in reality no more than a passive actuary or "recorder" of the objective situation. The subject then had to write down her estimates of the extension which happened to be most prominently attended to by her as "figure" in her visual field of the moment, as well as of other elements of the situation, shifting from one of five attitudes to another. The attitudes were b and p, b' and p' (see §§ IV and V), and one requiring simple distance judgments, d. The extensions in question, as well as their positions and distances from the subject, were afterward measured, computed, or otherwise objectively ascertained by the recorder.

Of the total sample of 174 situations, all the 93 with figural extensions perpendicular to the line of regard were then selected for further analysis, thus eliminating problems of distortion of form and its correction in perceptual shape constancy. When plotted logarithmically, for reasons explained above in connection with Experiment C as well as in the original monograph (1944,

pp. 6 and 8), bodily sizes B actually fall in a normal distribution (as may be gathered from fig. 8, but as is more clearly shown in fig. 1 of the original monograph). They range over as much as about five powers of 10, a range quite unheard of in the study of individual differences. This fact is noteworthy in view of the high correlations obtained for perceptual size constancy (see below).

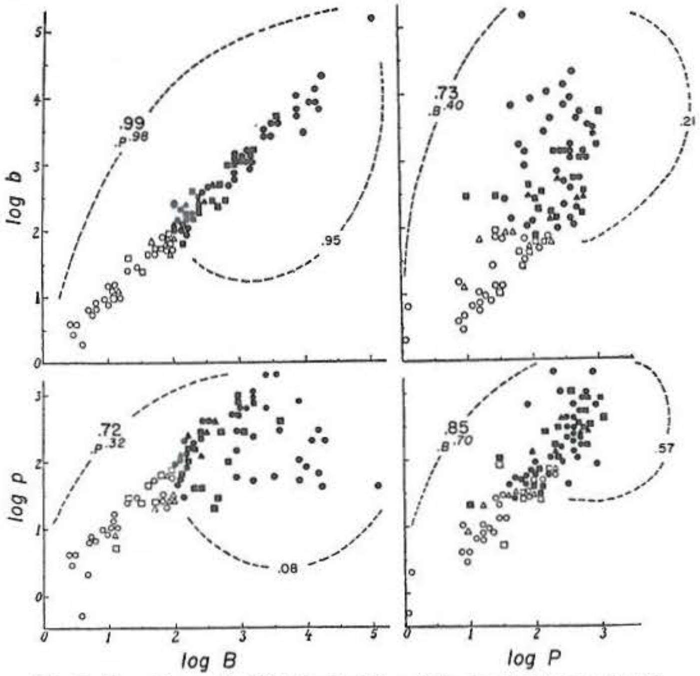


Fig. 9. (From Brunswik, 1944, fig. 3, with coefficients added from table 5.)

Stimulus-Response Correlations Representing Perceptual Size Constancy in One Subject.—Another intercombination of the data summarized in figure 8 by the large symbols and the arrowheads to allow more convenient inspection of perceptual achievement in terms of correlation rather than in terms of error. The ecological variables B and P are plotted horizontally, dependent response variables b and p vertically in a fourfold arrangement. The more fundamental "homonymous" combinations Bb and Pp appear in the upper left and lower right. Correlation coefficients for the total samples of 93 as well as for the 59 situations with sizes over 10 cm (the latter in small numerals on the right) are again based on logarithms and presented in a manner analogous to figure 8. Partial coefficients are in small *italics;* left subscripts preceded by a period mark indicate the variable partialed out.

There is a tendency toward curvilinearity in the relationship between B and P as plotted; this is obviously due to the limit of near vision at about ten inches, witness the fact that this trend is lost when small objects (under 10 cm) are excluded. For the sample of 59 objects or situations thus left, the ecological correlation between the logarithms of the measured bodily extensions and their retinal projections, designated in figure 8 as r^{log}_{BP}, is only .14 instead of .70 as in the full sample of 93, the (part-)causal relationship existing between B and P notwithstanding. (Coefficients from the numerical values of the variables involved were not computed.)

The discrepancy between statistical and causal dependency is in the present case readily dissolved by extending the scope of consideration to a third variable, distance, *D*. Together with *B* as a further part-cause of *P*, *D* strictly determines retinal size in accordance with the "laws" of geometrical optics. Following the more macroscopic, or "molar," branches of functional psychology, we are interested in the correct intuitive extrapolation of bodily size regardless of changes in distance. The variable *D*, *per se*, has therefore no place in such appraisals of size constancy as that derived below from figure 9, at least not so long as the achievement aspect is not superseded by an interest in particulars of functional mediation. The entire slant of studies of this kind is thus inherently correlation-statistical in the narrower sense of the word; or at least it becomes so as soon as an adequate, statistically representative approach is chosen. A schema of the entire problem pattern involved is given below in figure 10, in a manner analogous to figure 7; the relationship of the strict law, involving three ecological variables, to the imperfect ecological correlations involving two variables which may be raised to perfect correlations by partialing out the third variable,[17] is also shown.

Results in terms of the subject's responses—note again that only one subject is needed for this type of correlation analysis—are most readily revealed by the scattergrams and correlation coefficients shown in figure 9. Distal focusing, that is, a high degree of perceptual size constancy established by the organism, is indicated by the nearly perfect functional validity of *b* with respect to *B* as represented in the upper left scattergram. The logarithmically computed coefficient is about .99 for the total sample of 93 situations (.98 when *P* is partialed out), and .95 for the larger objects alone. There can be little doubt about the ecological statistical significance of these results.

More precisely, the first value is .969 uncorrected; and it is .988 when corrected for grouping the logarithms in intervals of .5 as indicated by the cells in figure 8, as are all other coefficients shown in figures 8 to 10 (Brunswik, 1944, p. 21).

Computed directly from the logarithms without grouping, and using three significant places of these logarithms, the correlation between *B* and *b* is .987. The corresponding value for $B \times b'$ is .994; and for $b \times b'$ we obtain .992. High as this latter correlation is, in line with the generally high correlations just cited, it nonetheless reveals an amount of independent variability of the naïve- and the critical-realistic attitudes which, though extremely limited, is about the same as is possessed by judgments given in these attitudes when compared with the measured distal variable, *B*. This fact, in connection with the discussion by the present writer (1944, pp. 14 f.) may be held against objections raised to the at least partially genuine perceptual (rather than intellectual or memory) character of the performance of our subject.

The correlation of *b* with *b'* at the same time supplies the closest approximation in the data of our subject to the problem of intra-individual observational reliability of the size constancy experiment, although it must not be forgotten that there is an important difference between the two attitudes which obliterates their equivalence, at least to a certain extent.

[17] It is in a vein analogous to the one just outlined that the present writer (1943) has questioned the value of the "nomothetic" search for strict laws in psychology proper. He is actually far from being a skeptic concerning the "existence" of generally valid laws, in psychology or otherwise, as he was misunderstood to be by Hull (1943*a*, p. 273). In a third paper, completing a "Symposium on Psychology and Scientific Method," Lewin (1943) joins Hull in espousing the nomothetic bias regardless of all the differences existing in other respects between these two outstanding theorists of psychology. See also Brunswik (1952).

In order to neglect as little as possible the aspect of populational generality of the results, the recorder was instructed to make, independently, estimates similar to those of the subject, and from the same position, before measuring the relevant features of the situation in question. Ample evidence is given in the present writer's monograph (1944) of the similarity of the results obtained from the two persons. Correlations computed from the original, ungrouped logarithmic data in addition to the evidence already published reveal agreements (inter-individual observational reliabilities) of .976 for *b*, and of .987 for *b'* between subject and recorder for the total sample of situations. (To compare with the functional validities and intra-individual reliabilities reported above for the main subject, the recorder's value for $B \times b$ is .993, for $B \times b'$ it is .989, and for $b \times b'$ it is .993.)

From beginners clinging to the concrete application of statistics to problems of individual differences as almost exclusively emphasized in classroom and text one is apt to hear—sometimes with a "know-better" air of supremacy—that correlations as high as those reported above "do not exist" in psychology. Granted that this is by and large correct for differential psychology, such high coefficients seem entirely possible in the present stimulus-response statistics. This holds especially when variability is as enormous as here, the ratio of the smallest to the largest extension being of the order of magnitude of 1 to 100,000 (see above). In consequence, even large errors of estimation may be absorbed into a correlation yielding an over-all coefficient of close to 1. Actually, the maximal errors occasionally committed by the subject with respect to *B* in the corresponding attitude *b* and thus incorporated in our highest coefficients are in the neighborhood of .5 logarithmic units, i.e., about 1 : 3 or even 1 : 4, or 200 to 300 per cent if expressed in terms of the estimate (see the horizontal components of some of the arrows in figure 8, extending over as much as one cell interval). These are errors everyone will admit to be rarely surpassed in daily life except for rather freakish instances which could likewise be absorbed if related to a large total of observations of which they are extreme cases. An even better illustration than figure 8 is the upper left-hand scattergram in figure 9 where the deviations from the diagonal—considerable if translated into numerical terms—will be found compatible with the extremely clear-cut over-all trend of the entire distribution. Small differences in coefficients may then become important; hence some of the high correlations were given here to three decimals.

The astonishing degree of perfection of the particular stabilization mechanism constituting perceptual size constancy has thus been demonstrated in its ecological generality with respect to the universe of situations from which our sample was drawn. The comparative inefficiency of performance in terms of photographic (retinal) relationships, even when a conscious effort in that direction is made—likewise anticipated in results like those described in § V/1—becomes obvious from an inspection of the two right-hand scattergrams in figure 9. Errors with respect to *P* run as high as 1:10 in the painter's attitude, *p* (see also the vertical components in fig. 8), and coefficients are lowered accordingly to between .85 and .57.

The point of discussion which actually supplied the impetus for the statistical survey of size constancy here discussed was the question of the ecological generality of the principle of perceptual compromise, referring to the slight underestimation of distant objects. This principle was mentioned above (§ V/1) in discussion of the size-constancy experiment. The question may be approached by an analysis of the errors as indicated by the direction and magnitude of the horizontal components of the small arrows issuing from each of the 93 large situation symbols in figure 8. As was shown quantitatively in the original monograph (Brunswik, 1944), the arrows starting from objects close to the subject—objects which thus have relatively large retinal projections—

indicate significantly more often an overestimation (or comparatively little underestimation) of bodily size than do those issuing from objects at large distances. The ecological applicability, in the sense defined in § VI/5, of the principle of perceptual compromise has thus been adequately demonstrated, as could never be done by a purely systematic experiment. This has been done at least for one—or rather two—certainly not atypical subjects. (For other, more traditional problems of perception likewise borne upon by the present study see below, § IX/2).

How practicable the general type of representative rather than systematic approach just outlined may prove to be, we leave open. The study here reported has certainly consumed an almost inordinate amount of working time, but the technique could probably be considerably simplified and standardized to a few representative situations. It must be added that representativeness of research design (representative variation and especially representative covariation) is most urgently called for—and will in turn prove most rewarding from a research point of view—in the successfully established stabilization mechanisms such as the thing constancies with their inherent ecological generality yet variable and multiple mediation (variable retinal image for constant bodily size combined with distance cues which in themselves are vicariously intersubstitutable, see § VIII), rather than in the relatively poor stimulus-response relationships found in the illusions or in social trait perception.

If such a program were carried out on a variety of functional topics, a subject (or patient) would then be described in his relationships to the world by a set of correlation coefficients (or other measures achieving the same end). That is to say, he would be psychologically characterized "in terms of objects" (ecological variables) he is capable of attaining (Brunswik, 1934, 1937) rather than just in terms of his responses, or in terms of relatively short-range achievements under specific conditions as in the classical experiment and test. His psychological portrait would thus emerge in terms of the stabilized, generalized object relationships established and maintained by him cognitively or in overt behavior, up to and including such wide-spanning covert distal adjustment features as, say, "social perceptual alertness and differentiation" vs. "social blindness," as may be approached through experiments of type D. From the coefficients describing these relationships one would, of course, not be able to predict correctness of orientation in the environment with certainty for any particular instance. But they definitely would give an over-all relative frequency ("probability") of adequate contact with vitally relevant variables, which may be a much more relevant type of information than certainty with respect to relatively insignificant instances or details.

VIII. Ecological Validity of Potential Cues and Their Utilization in Perception

Any fairly consistent rapport, be it intuitively perceptual or explicitly rational, with distal layers of the environment presupposes the existence of proximal sensory cues of some degree of ecological validity to serve as mediators of the relationship.

1. Ecological Analysis of Physiognomic Cues

Characteristically, attention to the ecological validity of perceptual cues was, until very recently, almost exclusively limited to the field of physiognomics, where the objects are persons and thus statistics have a foothold by virtue of individual-differences problems. A generally low degree of ecological validity of physiognomic cues—surprising as this may be to the layman—has been found in a large number of studies, summarized by Paterson (1930) and by Jones (1937).[18] Correlations of about .10 or .15 for such combinations as height with IQ are, in spite of being low, among the best established. Similar values have also been obtained in our Experiment D (§ VI/1 and fig. 6). (The height-weight ratio claimed by some authors to correlate about .3 with intelligence gave in our material correlations with the various measures and estimates of intelligence averaging under those reported in figure 6 for height alone.) Linking emotional states in social objects—i.e., *a*- rather than *u*-variables—with their external or peripheral-physiological "expressions" (in the narrower sense of the word) has been found somewhat more successful (and thus their social perception more valid, see § VII/2).

2. Ecological Analysis of Distance Criteria

Cues used in the perception of physical objects proper, as well as criteria of situational circumstances such as, especially, of distance, have in the past been approached primarily under the purely technological aspect of physiological mediation mechanisms. The problem of the statistical validity of such cues was explicitly raised by the present writer (Brunswik, 1934, pp. 49 f., 111 f., 116 f., 224; see also his summary in English, 1937) and further discussed in a joint article with Tolman (1935). In the statistical survey of size constancy summarized in the preceding section (Brunswik, 1944, pp. 9 ff., 42 ff.) it was decided to postpone investigation of the effectiveness of the distance cues involved for later occasions. It was emphasized, however, that an investigation of this type of mediational problem—broader than physiological technology—would have been quite possible if, say, the situations of the survey had been photographed from the place where the subject stood (thus establishing a duplicate of retinal configurations) and these records had been subjected to further scrutiny.

An analysis of this special kind, limited, however, to the problem of the ecological validity of distance cues, is undertaken in a forthcoming study by Stanford E. Seidner, who uses randomly selected pictures from a popular magazine and randomly selected places (objects, points), or pairs of places, within these pictures.

Preliminary results obtained by Seidner, as yet unpublished, suggest the following ecological validities, i.e., correlations of the real distances in the situations photographed—reconstructed from the pictures in terms of a crude five-point scale with a sufficient degree of reliability—with the actual location, color, etc. of their projections in the photographs:

About .6 for the cue of "vertical position," i.e., for the probability of greater real distance for objects appearing higher up on the picture.

[18] For some specific problems connected with details of facial and bodily geometry see also Cleeton and Knight (1924), as well as Laird's report of their study (1925).

About .4 (biserial) for "filling of space" (measured by the number of items, i.e., distinguishable steps, between the projection of two objects), more items between objects being associated with greater differences in real depth.

About .2 for "color" (i.e., on the achromatic pictures used, the local brightness of a spot), greater brightness increasing the chances of greater real distance—as would blue vs. red.

The relatively small and possibly biased samples of situations so far used by Seidner—75 instances from 12 pictures for each cue,—as well as the inherent inaccuracy in the determination of real distances from photographs alone and, finally, certain difficulties in precisely defining the variables, especially the retinal gradients, involved, make the numerical values listed subject to considerable revision. In figure 10 the obtained coefficients are given to two decimals; a question mark is added to indicate the tentative character of the results.

3. UTILIZATION OF CUES IN PERCEPTION

In view of all this, one is reminded of Thurstone's remark that perception is based on insufficient evidence, or of William James's saying that perception is of probable things. Looking at cues of limited trustworthiness under the aspect of biological adjustment rather than under that of physiological mediation—a policy with which such authors as Boring (1942, p. 303) seem also in sympathy,—their ecological validities, defining more specifically their inherent potentialities as representatives of a more distal factor, should be mirrored within the response system of the organism by the actual effectiveness of the cue in establishing a specific reaction focused on the more distal variable. This effectiveness, or degree of utilization (or impression value, or subjective weight, or response eliciting power) of a cue may be small or entirely absent, even in the face of considerable ecological validity; or it may surpass it, as it probably does in many of the social stereotypes formed in our culture, although its share of distal functional validity cannot consistently exceed its ecological validity. In a perceptually well-adjusted organism or species, however, the rank order of utilization in what may be called the "or-assembly"[19] of cues, or the "cue family hierarchy," should be the same as the order of their ecological validity. In consequence, the two aspects are best studied together. For social perception, this is illustrated in figure 6 for what in that field is called the expression problem (in the widest sense of the term) and the impression problem; see also figure 7.

A synthesis of the ecological with the utilization aspect of distance cues comparable to that presented in figure 6, with both of these aspects further subordinated to the related over-all achievement aspect of size constancy, is presented in figure 10. So far as cues are concerned, preliminary ecological validities were available, thanks to the above-mentioned efforts of Mr. Seidner. No representative data are available, however, for the utilization aspect of the distance cues proper. For this reason, findings of a pioneer laboratory study by Schriever (1925; see also Woodworth, 1938, pp. 664 f.) on

[19] This is a literal translation of the present writer's term "Oder-Verbindung" (1934, pp. 112, 191 ff.), designating the alternatives intersubstitutable and bound together—in the sense of the inclusive "either-and/or"—in the "vicarious functioning" (the term by Hunter; see Brunswik, 1943) of a unified stabilization mechanism. The term "cue family hierarchy" used above as an alternate expression has been chosen to stress the analogy to Hull's concept of "habit family hierarchy" (1934), introduced, simultaneously with and independently of the present writer's first publication on vicariousness of perceptual cues, in an effort to describe vicarious functioning in the field of overt action.

the rivalry of conflicting distance cues, using arbitrarily designed stereoscopic drawings, had to be substituted.

"Interception" (covering, overlapping of two objects, proximally perhaps best definable as absence of "good continuation" of contours) proved to be the most powerful under the conditions chosen by Schriever. This is well in line with the probably extreme ecological

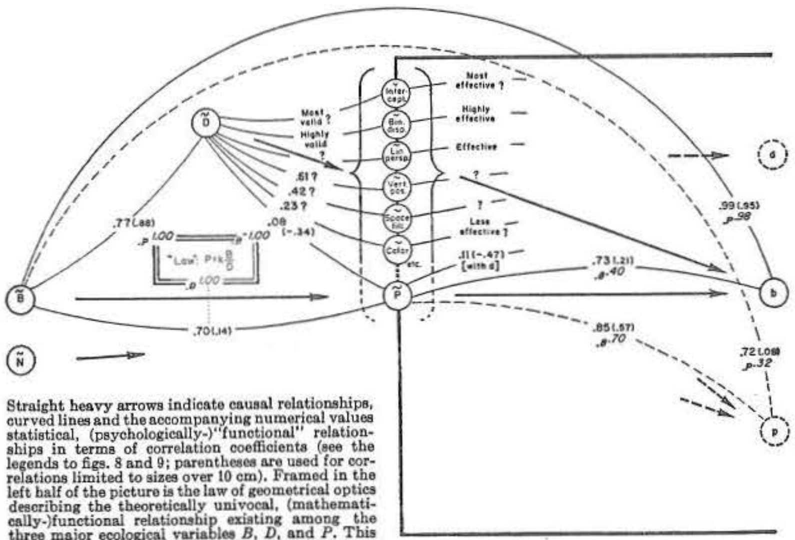
Distal stimulus variables and intra-ecological relationships. Correlation coefficients by Brunswik (1944), rearranged from fig. 8, above.

Hierarchy of proximal distance cues (including the direct retinal image, P)

Intra-ecological validities relative to actual depth, *D*.—Preliminary correlations by Seidner (see § VIII/2).

Utilization by the perceptual system (response effectiveness).—For depth cues proper, possibly nonrepresentative experimental results from Schriever (1925); for *P*, newly computed coefficients based on the original material of Brunswik (1944).

Estimates by a subject and functional validities (perceptual achievements) with respect to the proximal and distal stimulus variables, *P* and *B*.—Correlation coefficients by Brunswik (1944), rearranged from fig. 9, above.



Straight heavy arrows indicate causal relationships, curved lines and the accompanying numerical values statistical, (psychologically-)"functional" relationships in terms of correlation coefficients (see the legends to figs. 8 and 9; parentheses are used for correlations limited to sizes over 10 cm). Framed in the left half of the picture is the law of geometrical optics describing the theoretically univocal, (mathematically-)functional relationship existing among the three major ecological variables *B*, *D*, and *P*. This law is also reflected in their perfect partial intercorrelations of +1.00 or −1.00. For the meaning of other symbols and abbreviations see text and fig. 7.

Fig. 10. (Data from Brunswik; Seidner; Schriever.)

Schema of Perceptual Stabilization Mechanisms as Exemplified by Size Constancy, Including Their Ecological Foundation, with Special Emphasis on Distance Criteria.

validity of this cue. "Binocular disparity"—by no means the absolute monarch of distance cues that it is often assumed to be in classical depth psychology—turned out to be just about capable of matching the joint effect of "linear perspective" (converging of line systems toward a common origin, as in the photographic representation of railroad tracks) and of "shading" (distribution of light and shade, related to, but not to be confused with, Seidner's "color" cue as mentioned above). The fact that linear perspective is not highly response-effective may be attributed to its probably not being too highly valid as a cue since the geometrical patterns involved may also be caused by actual star- or trapezium-shaped objects in a frontal position.

The size of direct retinal object projection, *P*, may, as suggested by the ecological coefficients in the left-hand part of figure 10, in itself be considered a distance cue of some sort, though certainly one of very low and possibly of negative validity (.08 for our total sample of 93 objects, and −.34 when the sizes smaller than 10 cm which crowd the limit of near-vision are excluded; the latter coefficient is significant at the 1 per cent level for the remaining sample of 59 situations). As should be expected, the organismic effectiveness of P as a distance cue is likewise low (.11 and −.47, respectively) but of perhaps noteworthy numerical correspondence to its ecological validity just referred to.

IX. Ecological Overgeneralization of Experimental Results in the History of Psychology

Premature generalization of results of systematic experiments are numerous in psychological literature. Some of them are comparatively irrelevant; others have had consequences highly important either in a practical or a theoretical sense.

1. An example of overgeneralization in perceptual testing

One of the most prominent experiments on depth perception is represented by the frequently used device of adjusting two or three rods (or threads, etc.) to equal depth in a situation devoid, or nearly devoid, of depth cues except the binocular mechanism (survey of the literature in Woodworth, 1938, pp. 665–674). Experiments of this kind were given as a test to inductees during the Second World War as part of the routine medical examination. As seems to be agreed upon, they did not lead to any appreciable correlation with the actual performance of flyers in situations discriminatory of orientation in depth, such as landing an airplane. Obviously, validity should never have been expected of this test since the kinds of criteria available in the classical experiment upon which it is based, on the one hand, and in the practical situation referred to, on the other, have almost nothing in common. Only occasionally, as in the work of Gibson and his associates (final publication, 1947), and too late for extensive practical use, has the necessity of adequate representativeness and closeness to life of the situations and cues used in the war testing of depth perception been recognized.

In testing in general, the first major awareness of the necessity of presenting representative tasks can perhaps be found in Binet, designer of intelligence tests at the turn of the century. Applied psychology has also tried to make tests representative or "analogous" to the total pattern of activities involved, but has been more inclined to do so by an "understanding" type of analysis than by methodical sampling. A set of experiments highly representative of the social and emotional aspects of the anticipated tasks to be performed has been developed during the war in the course of an official "assessment" program (see Murray and MacKinnon, 1946). As has been pointed out repeatedly, however, intrinsic similarity between test and task must be superseded by considerations of statistical validity through which ultimately preference may be given to a seemingly unrelated test.

In a technical sense, the notion of "sampling of tests" originates, as Professor Hotelling informs me, with him (1933); but he only has in mind sampling from a battery of tests the members of which may have been constructed after arbitrary or speculative principles. There can be no doubt that considerations such as those advanced by Thurstone, dealing with the question of whether or not several different experiments or test batteries yield the same set of identifiable primary factors, move under the inherent pressure of the long-neglected aspect of ecological representativeness of task situations and tests (see Thurstone, 1940, especially p. 195 and the subsequent discussion with Thomson on the fundamental question of the functional "reality" of factors, 201 ff.). The same holds for Thomson's own

anticipation of a sampling of "the whole pool of a mind" (1939), and possibly also for Lindquist's not further elaborated hint toward a "random selection of schools" from which to sample children for an experiment (1940).

One has the impression, however, that the authors mentioned are not fully aware of the broadness of the underlying issues involved in the concrete steps they have taken so far. Thus, the synaptic junction with whatever complementary growth may be forthcoming from revisions of general experimental methodology as suggested in the present paper is probably still a long way off.

A fully adequate handling of the question of the representativeness of a certain test with respect to a certain extrinsic criterion presupposes, as is generally accepted, a shift of the empirical validation from other tests, or comparatively irrelevant behavior (such as, e.g., scholarship) to the direct behavioral issue or clinical symptom involved. A good example is Strong's (1943) vocational aptitude testing as validated against actual success in life. Assessment programs such as that mentioned above seem to be working toward a similar goal as is the case with other recent developments in psychological testing.

2. Premature application of special findings within academic psychology: vertical illusion and Gottschaldt experiment

Since statistical studies of stimulus-response relationships, such as the monograph on size constancy described in § VII/3 (Brunswik, 1944) can be evaluated in many different directions, this investigation has been used to throw light, on a representative basis, upon some problems not directly related to the thing constancies. The problems are the so-called vertical illusion and the Weber law, both old standbys in experimental psychology.

Allegedly, a vertical is overestimated as compared with a horizontal of equal length. Examples to illustrate the point are sometimes glaringly unique and full of secondary sources of illusions, such as the well-known "silk hat" pattern dating back into the nineteenth century but still revived in some contemporary treatises such as that by Luckiesh (1922, pp. 46 f.). Systematic experiments were limited to the square as well as L- and T-patterns, variously inverted and tilted, and related arrangements of geometrical lines. The inverted T suggested by Titchener and involving the added illusion of the bisected line has, in various modifications, become a favorite in textbooks, not excluding one by the present writer (1935, fig. 110). For a history of the problem see Finger and Spelt (1947); see also Würsten (1946).

However, comparisons undertaken under a variety of aspects within the material of the size-constancy survey (§ IV/D and table 3 of Brunswik, 1944) have not borne out the ecological applicability of this illusion. There is even a slight trend in the opposite direction, although statistical significance is hampered by the relatively small number of vertical extensions, 25, as contrasted with 57 horizontals (see also above, fig. 8).

The Weber law is the first genuinely psychological quantitative law, and the backbone of Fechner's famous volumes on psychophysics (1860). In contradistinction to the vertical illusion, this law has been confirmed by our results in its ecological generality for length discrimination. Graphically, it may be read from figure 8 of this paper by establishing the approximately equal average length of arrows for small versus large sizes, as well as small versus large projections (for details see § IV/E of the monograph cited).

Some experiments are especially designed to prove—or, at least allegedly, to challenge—pivotal points of discussion upon which the fate of an entire psychological movement may hinge.

An outstanding example is the much-cited experiment by Gottschaldt (1926, 1929) with "masked figures" (one of these is reproduced in Woodworth, 1938, p. 640). In spite of being "learned" in hundreds of presentations, the simple line drawings involved failed, by and large, to be perceived as such when embedded within larger designs strongly held together by such "autochthonous" factors as symmetry, completeness, and "good continuation." The conclusion, commonly drawn in Gestalt psychology, that "experience" is of negligible or at least subordinate influence upon perceptual organization (see Koffka, 1935) seems unjustified in its generality in the light of the restricted type of material and the unusual strength and mutual support of the integrating factors rivaling experience. The experiment proves nothing but the possibility that experience may not be an overpowering factor in a specific situation. Responses to the well-known Rorschach ink blots, not to mention systematic experiments dealing with related topics, including the memory for form, have shown beyond doubt that under many conditions it is the familiar rather than the geometrically clear-cut or "*prägnant*" that will stand out in perceptual organization.

3. LIMITS OF GENERALIZABILITY OF EXPERIMENTAL EVIDENCE

What, then, are the possibilities of applying results of an experiment to new situations by means of generalizing inference? The question can be answered in the same general way as it is customarily answered in the statistics of individual differences. Everyone knows that encountering, say, a wife who is taller than her husband (or, to come back to an example cited above, a railroad president who looks like a criminal) does not justify the inference that wives (or railroad presidents) are always, or are overwhelmingly, taller than their husbands (or indistinguishable from criminals). What the instances mentioned do demonstrate, however, is the fact that it is possible for a wife to be taller than her husband (or that not all railroad presidents can be told from criminals). They thus enable us to make minimum statements of the form "at least in certain individual cases . . ."

Experiments in the biological and social sciences are often formally analogous to the instances referred to above, by virtue of the fact that they do demonstrate a mere possibility; this time, however, specification is made by adding "at least under certain conditions . . . ," just as in the discussion (see above) of the Gottschaldt experiment. Experimental results are sometimes contingent upon situational conditions which may not become known unless further investigations effecting a separation of variables are undertaken. Examples of conditionally delimiting possibilities revealed in a systematic experiment are those given, say, by a staying alive (or of not being able to live) on a pure meat diet, or by a having of visual sensations on the basis of retinal stimuli other than light, or by an average differential threshold of 1/200 for weight discrimination (under certain conditions of kinesthetic comparison), or by repression being demonstrable (or not demonstrated) in a certain memory experiment, etc.

These examples differ from the individual instances cited at the beginning of this subsection by the fact that they are usually established with a certain degree of populational generality. In this case, typical of the systematic experiment, evidence definitely goes beyond "anecdotal" atoms or protocols of knowledge which would involve *one* individual in *one* situation only. Ecologically, however, they are true in their own right only; they remain ecologically incidental as long as the generality of the functional principles involved is not technically established by studying them in a representative manner. Thus, experiments using the customary patterns of the vertical illusion referred

to above have, at least in human beings, quite consistently revealed a populationally generalizable overestimation of the particular verticals. However, no matter how many subjects were used, in an ecological sense this is minimal material allowing no more than description of the particular paradigms chosen; this holds the more as these paradigms canvass an ecologically highly restricted type of stimuli or situations.

Quite often the demonstration of a mere possibility as outlined above is all that is necessary and desired of a piece of research, and may be fully sufficient to establish tentatively a principle for purposes of further verification and thus to stimulate further research; in all cases of this kind the systematic experiment is in place and may save the burdens that would go with a proof of ecological generality. In other cases, a systematic experiment may serve to exclude certain trivial factors from the explanation of a phenomenon.

On the other hand, the most striking shortcomings in the generalizability of results of systematically rather than representatively conducted experiments are given when it comes to a quantitative estimate of the relative contribution of competing factors in functional adjustment to the environment. Such interplay of forces is exemplified in the present paper by the rivalry of length and area, or of bodily size and retinal size, or of the various physiognomic cues or traits, as polar variables in a process of perceptual compromise or stabilization, as the case may be. This is also the type of problem to which factorial design and analysis of variance has been tailored; however, generalizability of results concerning the relative weights of the variables involved must remain limited unless at least the range, but better also the distribution of the "levels of strength" employed for each variable, has been made representative of a carefully defined universe of conditions.

X. Development toward Greater Representativeness in Experiments on Learning

1. LEARNING OF PROBABLE CONNECTIONS

There is one example of the interwovenness of classical and representative elements in modern experimentation which has special relevance to one of the main topics of the present Symposium: probability. The short tradition of this type of research in psychology makes it understandable that work in this direction has not yet outgrown the classical stage in most of its methodological aspects; like the thing constancies, however, the topic itself has the earmarks of representativeness.

In experiments of this kind on overt behavior in animals (reviewed by Hilgard and Marquis, 1940), the old black-white technique of "right" vs. "wrong" choices or turns in a maze—resembling the likewise in many ways nonrepresentative current code for movie stories—has, first, been replaced by "partial reinforcement" (taking place only every n-th trial and thus removing the previous inseparability of "correct" choice and reward), as in experiments by Pavlov, Skinner, Humphreys (1939), and others.

A further and probably more radical departure from classical univocality consists in letting the animal discriminate between higher and lesser degrees of probability, establishing what amounts to a "psychophysics of probability" (Brunswik, 1939; also reviewed in Hilgard and Marquis, 1940). The threshold for probability in rats, using a short series in a simple T-maze, was found to be between 67 per cent vs. 33 per cent, and 75 per cent vs. 25 per cent, the percentages indicating the relative frequencies of reward on the two sides of the maze.

In the field of perception, the learning of new cues for illumination and other situational circumstances or associations, sometimes tested by the establishment of certain illusions, has been studied by Fieandt (1936), Brunswik (1938), and Max M. Levin (1943, 1946).

In passing, it is interesting to note that Levin's more recent study (1946), in an artificially interlocked design confronting relative frequency regardless of total amount (weight in the particular case), with amount regardless of frequency, has found *amount, not frequency*, the crucial factor. But this challenging of the frequency principle—the dominant principle in the theory of association and conditioning—may of course again, coming from a systematic experiment in a specific field, not be ecologically generalizable. This holds although even a single situational instance—when populationally generalizable, as in Levin's study—cannot be passed over lightly when the validity of a fundamental theoretical principle is at stake.

In a quite different context, the frequency principle may also be questioned in its universal applicability by the existence of the "gambler's fallacy," i.e., the anticipation by the typical subject of that one of two alternatives which has occurred *less* frequently in an immediately preceding series of coin throwings or presentations. The occurrence of this fallacy has recently been demonstrated by Jarvik (1946) in experiments dealing with the serial anticipation of alternatives of different degrees of probability (relative frequency).

Increasing attention is also being paid by the modern formalizer of the theory of behavior, Hull (1943*b*, pp. 374 ff.), to what according to him may in time become a "calculus of adaptive probability." From outside of psychology, the strongest impetus and support for an emphasis on probability in psychology comes from Reichenbach (1938), with his conception of the "wager" character of behavior in the framework of his "probabilistic empiricism." This philosophical system is in certain respects akin to the "probabilistic functionalism" that may be suggested for psychology on the basis of the limited character of ecological and functional validities to which attention has been directed in the present paper.

2. CHANGES IN THE METHODOLOGY OF LEARNING EXPERIMENTS IN GENERAL

Developments similar to those discussed in this paper for the field of perception may also be observed in the field of learning, although perhaps with some temporal lag possibly connected with the fact that Ebbinghaus' pioneer study on memory (1885) came twenty-five years after Fechner's psychophysics. Ebbinghaus used nonsense syllables to eliminate, or to hold constant, the disturbing factor of meaning, quite in line with the philosophy according to which a black cloth is used to surround the two halves of the Galton bar in our Experiment A. Many later experiments (see Woodworth, 1938, pp. 205 ff., as well as the last two chapters of his book), and most recently those by Katona (1940), have made inroads into the exploration of apparently vast transfer,

stabilization, and generalization effects, such as the "learning of principles." Some of these effects seem quite comparable to those of the perceptual thing constancies.

There is some awareness of the problems of situational generalizability in classical memory research, and attempts at a synopsis under certain comprehensive abstract headings. As good an example as any is the assertion of the general superiority of passive "recognition" over active "recall" (see Woodworth, 1938, pp. 47 f.). This superiority is usually specified as being roughly in the ratio of two to one, sometimes three to one. Characteristically, however, the material is brought forward in what has been called (§ IV/1) a process of canvassing or of scattered exemplification (in the present case involving syllables, words, and proverbs) rather than in the form of true sampling; and the ostensibly major crux in the applicability of the results in question, namely, the difficulty of "recognition" when new items are used which closely resemble the old, is introduced by means of far-off, arbitrary materials (colors and geometric designs).

As a likewise somewhat retarded analogue to the increase in emphasis on the central factor of attitude in perception since the turn of the century, motivation has in the last few decades come to play an increasing role as an actual variable—rather than just as an indirectly controlled constant—in a great variety of learning experiments.

In the special field of animal maze learning, Watson's classical behaviorism of the 1910's and the 1920's with its search for "the" locus of learning—thought to be sensory or motor, but in any event peripherally channeled and focused—has given way to a recognition of the generality, and thus of the central character, of the learning mechanism. This progress was carried by Lashley and his group, and by the "molar" behaviorism of Tolman (1932) and his collaborators at the University of California, such as D. A. Macfarlane, Krechevsky, and Honzik, as well as by many others.[20]

Turning to formal methodology, the successive omission of sense departments or cues used in some of these experiments, in contrast to successive accumulation as practiced by the early behaviorists, revealed the intersubstitutability of cue systems quite in analogy to what was found by Holaday for perceptual distance cues with the use of the same technique (see § V/1, 2).

Even more in line with the natural conditions of the interplay of forces—encountered by the researcher in the perceptual constancies more or less automatically—is the method of successively altering or confusing the cues rather than successively eliminating them (see Woodworth, 1938, pp. 130 f.). In this way, stabilized achievement in the face of actual disturbances on the part of certain factors or cues, rather than just the dispensability of certain cues in the establishment of a certain achieved connection, becomes obvious.

Methodology in the field of learning has not yet reached the stage of full ecological representativeness as exemplified in the present paper by Experiment D on social perception as well as by the statistical survey of size constancy and by the ecological and utilization analysis of distance cues. The

[20] An excellent account of these developments is given by Woodworth (1938, first half of chap. vi, pp. 126–140). The California publications have in part appeared as Vols. 4 and 6 of the University of California Publications in Psychology, 1928–1934. Concerning the formal similarity of the situation in the fields of perception and of learning see Tolman and Brunswik (1935). See also Brunswik (1946*b*).

latter might in the future be paralleled, in the region of overt behavior, by a representative analysis of means (habits) in relationship to ends, both with respect to ecological validities and the hierarchy of utilization.

Summary

A sequence of four experiments, involving threshold, illusion (Gestalt dynamics), constancy of apparent size, and social perception of intelligence and personality traits under conditions of restricted contact, is chosen to cover a century of methodological progress in experimental psychology. In particular, the series is used to illustrate the changing over—under way on a broad front but still to be rendered explicitly acceptable—from a "classical" or at least "systematic," in the last analysis autocratic style of laboratory research to a more "representative" type of experiment that borrows heavily from the spirit of statistical surveying.

In this process, the practically exclusive hold of the spearhead, "individual differences," on representative sampling in psychology is being challenged, and representative sampling is extended from the subjects to the objects, from the individuals to the stimulus situations and tests. Thus the traditional double standard with respect to the status of organismic (and predominantly statistical) and stimulus (and predominantly systematic experimental) variables is being gradually removed, and psychology is conceived of as a fundamentally statistical discipline throughout its entire domain, with "functional validity" taking its place alongside traditional test validity. It then also becomes possible to ascertain the "ecological generality," or "applicability," of a result along with its counterpart, populational generality or reliability in the customary subject-centered meaning of the term as studied in the familiar context of differential psychology.

An example of the fusion thereby inaugurated—bringing to convergence European academic with Anglo-American statistical tradition—is also given, and fallacies, inconsistencies, and emotional resistances encountered are discussed.

Brief reference is also made to comparable developments in the field of learning which appears to lag behind the older and further-matured pioneer field of perception with respect to general methodological orientation.

PART TWO

PERCEPTION

THE ECOLOGICAL GENERALITY OF ITS DISTAL AIM

II

XI. Visual, Auditory, and Tactile-Kinesthetic Constancies: Canvassing by Systematic Experiment as Ecological Spot Sampling

The basic fact of perception is distal focusing. No orientation in an organized "world" would be possible without it. Distal focusing is the result of an ecological generalization process on the part of the responding organism. The generalization takes place over the range of the concrete variants in the proximal mediation patterns of the distal variable. The presence of such a generalization, together with some characteristic imperfections, was demonstrated for the visual size variable with the use of representative design in one subject (§ VII/3).

The question now arises whether this ecological generalization within one variable is further generalizable to other distal variables or other sense departments so that it may be regarded as an ecologically universal trait of an individual or species.

It may further be asked how slight a degree of encouragement will suffice to bring out the trend toward further penetration in depth from overt distal to covert distal types of objects, a trend which we have observed under favorable conditions in social perception (§ VI).

In all these contexts a distal "aim" will have to be recognized, even when full or approximate distal attainment is wanting, so long as the presence of a tendency away from a coördination of the response with the most confined proximal variables and toward a special positive distal instrumentality can be demonstrated. The subsidiaries of distal achievement that reside in relatively fixed or changing ecological validities of cues and in a correspondingly flexible machinery of cue utilization are part of this total picture. To unfold it by giving as great a variety as possible of concrete illustrations of the problems involved and of its variants is the purpose of the second part of this book.

One of the variants of distal focusing is given by the perceptual thing-constancies as we have introduced them in § V. Traditionally, thing-constancy research has centered about three subvarieties: size constancy with distance variant, shape constancy with tilt variant, and color (or whiteness) constancy with illumination variant.

Uexküll and Kriszat (1934) have vividly described what it would be like to live without these constancies. Figure 11 shows at the right the subjective or behavioral environment of a sea urchin, a lower animal capable of isolated response to pressure and chemical stimulation only. This implies an absence of three-dimensionality of space, and of color constancy. In consequence a cloud, a boat, and the real enemy, a fish, must be responded to in the same manner. This leads either to wasteful motion or to excessive exposure to danger. The authors bemoan the fact that in the drawing all three objects had to be represented in space while in reality their shadows act uniformly "like a touch with soft cotton gently sweeping the light-sensitive skin."

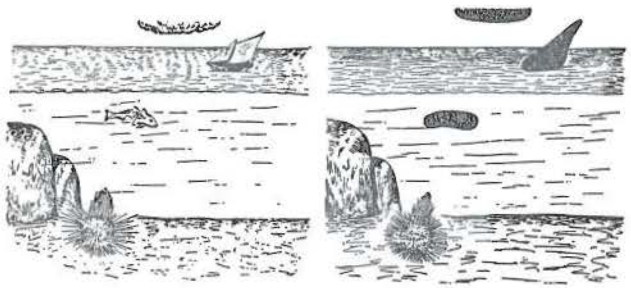


Fig. 11. (From Uexküll and Kriszat, 1934.)
Ecology vs. Behavioral Environment of Sea Urchin.

Under such general circumstances there can at best be no more than what Gibson (1950) calls the "visual field," and never a "visual world." There can be no orientation or manipulation as higher animals know it.

At the left of figure 11 is shown the measured geographic environment with all its relevant differentiation; thing constancy is nothing but the mechanism that makes the behavioral environment conform to the geographic environment to a considerable extent, especially in its biologically more relevant distal aspects, thus making higher life possible.

An important problem of functional psychology is the question as to exactly how perfect and how pervasive the mechanism of distal focusing is. As we have demonstrated in § VII/3 for the case of size constancy, the final answer about the degree of perfection must and can be given by representative design only. Here we will first turn to a more complete discussion of relevant systematic experiments of the relatively recent, more life-like kind and see how much they have been able to anticipate the final outcome and how they may be interpreted as parts of a representative design. In the remaining two subsections of the present chapter we will turn to expansions of the constancy problem to other sense departments. In § XII will follow what we will call the "extended constancy problem." Together they will give us an at least tentative answer to the problem of the pervasiveness of distal focusing. These considerations will also give us an occasion to elaborate on the systematic-representative hybrid techniques of "canvassing" which we have so far only briefly touched upon at the conclusion of Part One (p. 57).

1. CLOSE-TO-LIFE SYSTEMATIC STUDIES IN SIZE CONSTANCY

Of the relatively life-like type of systematic experiments in size constancy, that is, those which proceed from a situation with a fairly "normal" array of depth cues to more impoverished situations in a technique of "successive omission," we have so far discussed only the earliest, that by Holaday (1933; see above, p. 24 ff.). A more complete picture of the merits or demerits of this representatively tinted kind of systematic approach is obtained by comparing notes with some subsequent experiments which have proceeded by essentially the same functionalistic scheme. These are the studies by Holway and Boring (1941), by Gibson (1947), and by Joynson (1949). In each of these cases there is, superimposed upon an at least allegedly exemplary or typical core situation, a greater or lesser number of systematic variations, so that the design is still fundamentally systematic at least in its ramifications.

In figure 12 are shown some additional data from the study by Holaday. They deal with the dependence of size constancy upon the distances employed. The numerical constancy ratios that appear in Holaday's table 2 (1933, p. 463) are used in plotting, but the constancy ratios that are written in our figure 12 are the result of a logarithmic recomputation. All data are based on "normal" binocular viewing in the naïve-realistic attitude *b*, by ten subjects, and with the use of cubes (Standard = 7 cm) placed on the floor of a large, empty hall, the subject's head being slightly raised above the floor. In addition to the two distance combinations that have also been used in combination with the successive omission of cues and that therefore have previously appeared in our table 1 (framed in fig. 12) there are eight further distance combinations, and the distance ratios range to an extreme of 1:16.

For the given conditions and range of distances (between ½ and 8 m) the absolute magnitude of the constant error, given by the ordinate beyond the 7-cm line, tends to increase with increasing absolute distances as well as with increasing distance ratios. Correspondingly, there is a decrease of constancy ratios as we proceed from left to right within each of the charts for the various distance ratios. Constancy ratios range from .95 down to .71.

A cross-comparison of the first values in the four charts reveals that there is no simple decrease when the position of the Standard is held constant, however; rather, there is perhaps an "orthoscopic" distance at about 2 to 4 m where objects, at least of the kind and size employed by Holaday, appear most realistically "as they are" (*c* reaches about .95); Koffka (1935) speaks of the frontal orientation as an orthoscopic plane of rotation, and Katz's (1911) "normal" illumination conditions (see above, p. 5) supposedly yield orthoscopic color experiences.

Using distances up to ten times as large as Holaday—10 feet to 120 feet—Holway and Boring (1941) obtained from their five subjects almost ideal constancy even in monocular vision while binocular vision yielded slight overconstancy (fig. 13, top). Holway and Boring's conditions were unusually favorable for depth perception; the experiment was set up in corridors which, even though relatively dark, provided certain "constellations of light images" such as reflections from the waxed floor which must have exerted a rather coercive influence via the cue of linear perspective. Suspecting a space error, the authors have nonetheless corrected their results (by 10° rotation), as shown in the bottom part of figure 13. Binocular constancy then would be, roughly, about .9 or slightly over.

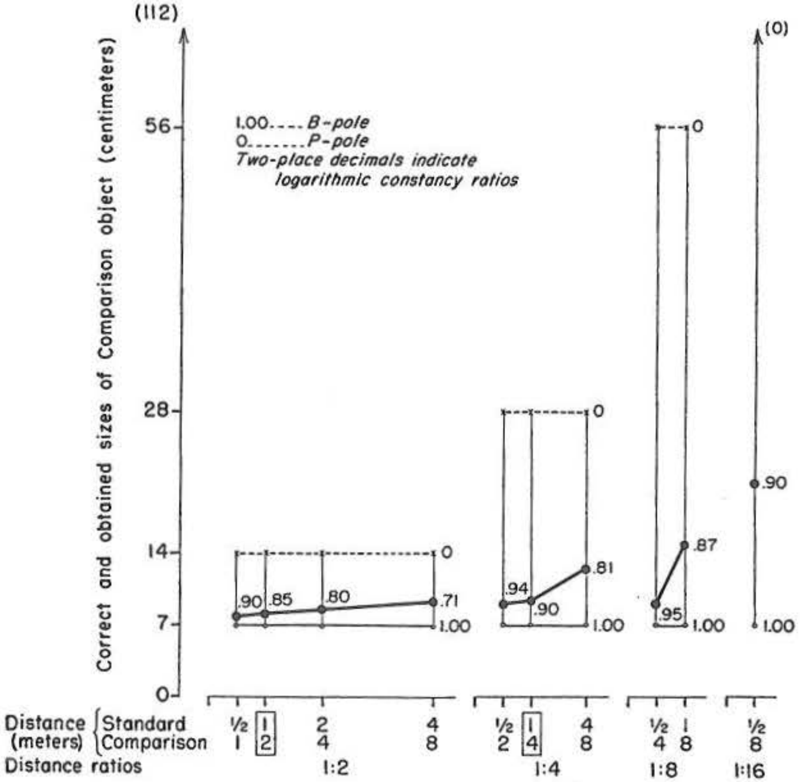


Fig. 12. (From data by Holaday, 1933.)

Dependence of Size Constancy upon Absolute Distances and Distance Ratios.—The framed distance combinations are those upon which the results in table 1 (p. 25) are based.

Holway and Boring have also followed Holaday's successive omission technique, notably in using "a reduction tunnel, a long black tube which would eliminate the perception of reflected light from the surfaces of the corridor." As in the case of Holaday's monocular darkroom series with fixed head (our table 1 reports c = .15 for this condition), Holway and Boring find a far-reaching breakdown of size constancy under these conditions.

Nearly perfect constancy, with some overconstancy, was also exhibited in the study by Gibson; in this experiment "size perception was studied up to

the point where the size showed signs of beginning to disappear optically" (1947, p. 210)—that is, up to 448 yards and occasionally even 784 yards—with standard sticks of from 63 to 75 inches at 14 yards (fig. 14). The example from the same study which Gibson discusses at a later occasion (1950, p. 186) shows the usual compromise tendency (underconstancy), however. The experiment had been conducted in a level field of cultivated land (as shown in Gibson, 1950, fig. 74, p. 184 f.). The texture gradient must have

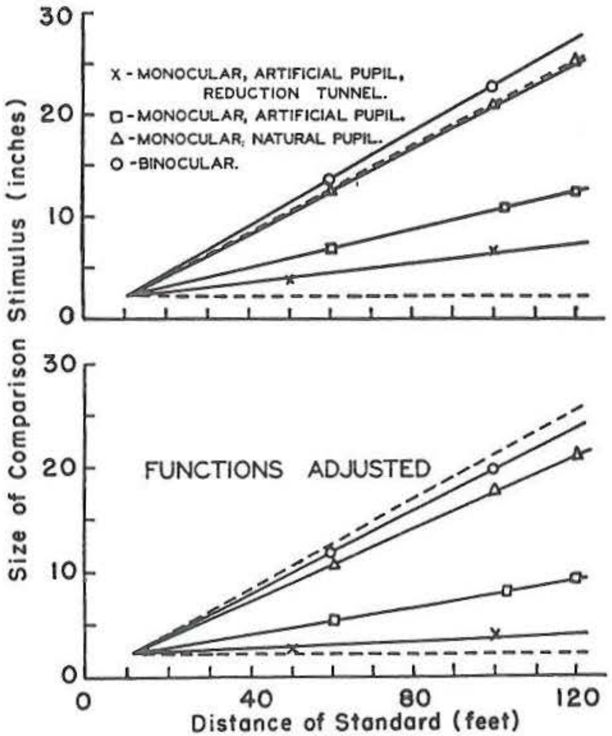


Fig. 13. (After Holway and Boring, 1941.)

Size Constancy in a Long Corridor.—The distance of the Comparison stimulus is constant at 10 feet. In each graph the upper broken line represents ideal constancy, the broken horizontal line ideal photographic vision.

furnished an almost equally inescapable depth cue as Holway and Boring's corridor, even though there were no furrows to give linear perspective.

Conditions in which one or more cues are as much tied to distance as in parts of Holway and Boring's or in Gibson's experiment are not representative but constitute unrealistic idealizations. In effect, they approximate classical tied-variable designs (in the sense of § II) as were exemplified in Part One by the Galton bar (§ III). Size constancy is almost automatic in such cases (or in their approximations at hand), which means that it is not really being tested. What is tested is merely the utilization of the cue or cues in

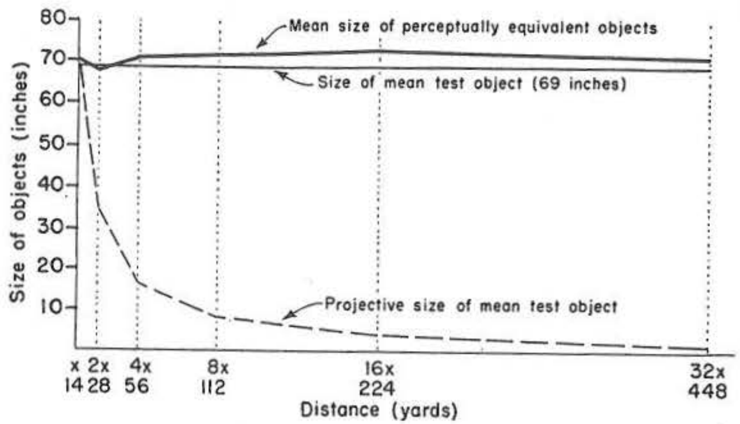


Fig. 14. (After Gibson, 1947.)
Size Constancy up to the "Vanishing Point" in an Outdoor Situation.

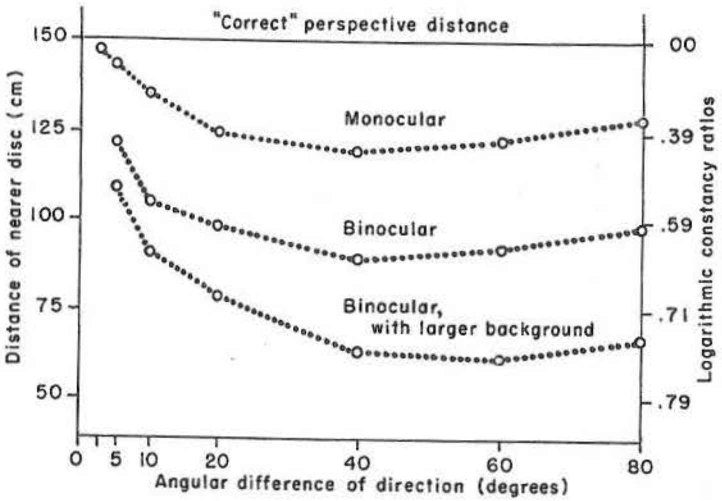


Fig. 15. (After Joynson, 1949.)
Size Constancy and Angle of Separation between Standard and Comparison, in Analytic Attitude.—In this experiment only two objects of different size were used, and the distance was made variable to obtain response equivalence.

question, and this is a mediational rather than a functional problem. Perfect constancy under such circumstances is no more than a tautology.

It is significant from the methodological point of view that in two of the studies the most favorable cue-conditions yielded on the average slight over- rather than undercompensation for distance ("overconstancy"), and that the over-all level of "compromise" varies considerably from study to study. Contrary to what Holway and Boring impute (1941, p. 34), this writer finds occasional or conditional overcompensation for distance not only compatible, but entirely in accord with expectations entertained by the functional point of view, a point of view mindful of the adjustive value of overcompensation elsewhere in nature also. Even in our representative design an instance of overconstancy corresponding to a c ratio of 1.9 could be reported (Brunswik, 1944, p. 30); but ratios slightly under 1.00 appear to be the ecologically most typical.

Holway and Boring had used two perpendicular corridors at the intersection of which the observer sat. In order to ascertain whether a decrease in the angle between Standard and Comparison from 90° to close to 0 would tend to diminish constancy, Joynson (1949) undertook the experiment described in figure 15. An analytic attitude was employed, and dotted lines were therefore used in drawing the curves. Logarithmic constancies (Thouless ratios) are nonetheless as high as in the middle .70's under reasonably favorable conditions, especially at angles 40° and 60° which still permit some sort of simultaneous comparison. But they drop sharply as the objects come closer to intercepting each other. Although the smallest angular separation is as large as 2½° and no actual overlap occurs, an imaginary extrapolation of the curves to 0° points to perfect photographic vision.

An intermediate step between the close-to-life systematic designs described in the present chapter and the fully representative survey reported in § VII/3 is given by an earlier study by this writer (1940) in which objects possessing predetermined, systematically varied sizes were scattered, at predetermined distances, throughout a predetermined room (see fig. 16). At least a certain measure of randomness in the background cues was thus achieved. The correlation coefficient was used for the first time instead of the constancy ratio to measure achievement, but all this constitutes no more than a superficial departure from systematic design. The results, also presented in figure 16 and summarized in figure 17, happen to be quite similar to those of the later representative study (summarized here in figs. 8, 9, and 10). We must remain aware of the fact, however, that the coefficients in figure 17 cannot be legitimately generalized beyond the artificial laboratory ecology incidental to the experiment including the particular size and distance ranges chosen within it.

Lumping all conditions together, the functional studies of size constancy just discussed bear out Hering's old assumption of "approximate size constancy," according to which more distant objects tend to be slightly underestimated in the direction of a compromise with photographic size. In view of the fact that our representative design (§ VII/3) has confirmed this over-

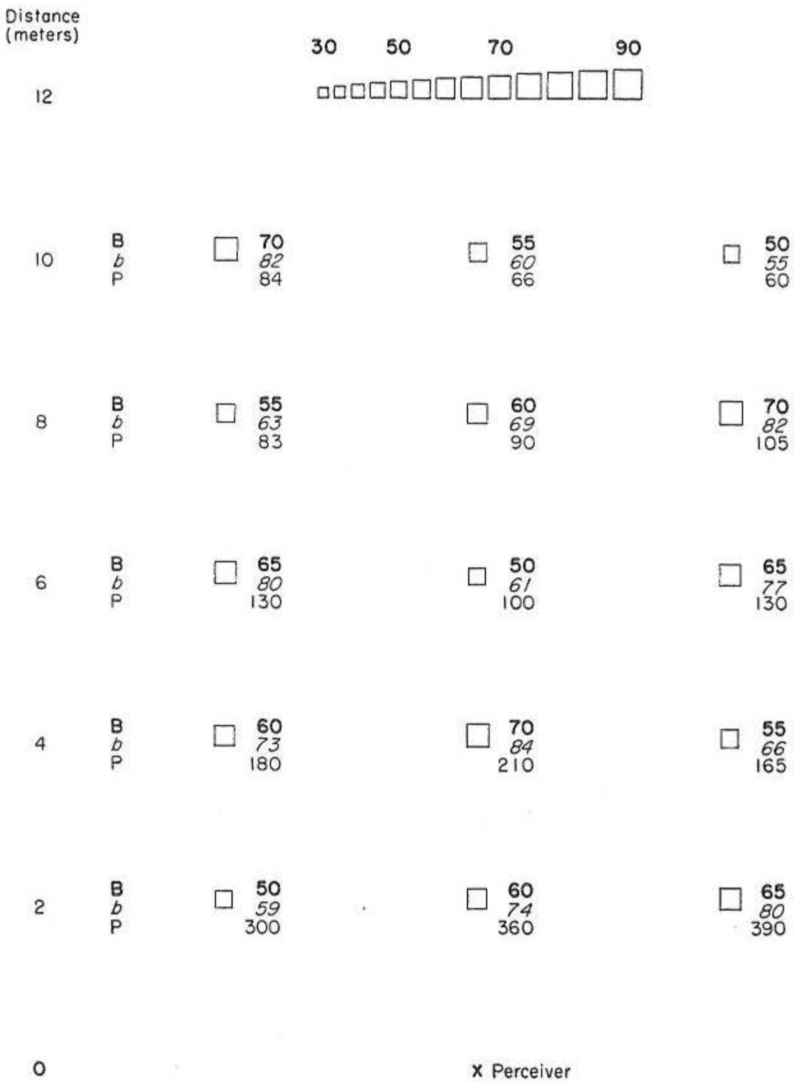


Fig. 16. (After Brunswik, 1940.)

Schema of Systematic Design in Size Constancy for Use with the Correlation Coefficient.—Front-to-back distances are foreshortened in ratio 1:4 when compared with lateral distances and the sizes of the cubes.

all result, the traditional trust in the informal "canvassing" inherent in the totality of these experiments would appear to be reinforced. Under the stricter scrutiny of representative design, however, all the systematic experiments listed must be viewed as ecologically incidental single cases or "instances." The artificial elaboration of these instances in systematic design leaves a large portion of the core elements untouched; their size-depth patterns are fixed in advance. More importantly, the particular strengths of the cues and their validities are rendered incidental and thus may be distorted as in any individual selection from the ecology. This is the more true

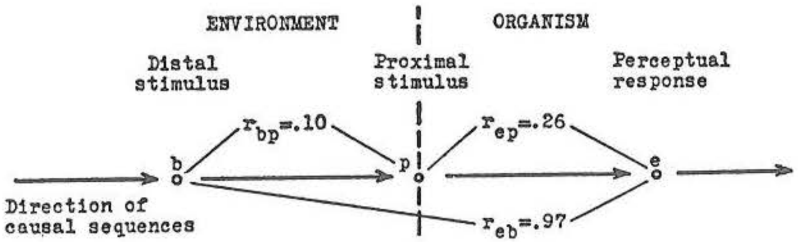


Fig. 17. (Redrawn after Brunswik, 1940.)

Correlations Between Distal, Proximal, and Response Variables in Systematic Size Constancy Design Shown in Fig. 16.

as in this type of experiment—including the one by Holaday in which this writer has had a hand—there never has been anywhere near complete realization of all the intricate variabilities that adequate representativeness would entail. Laboratory experiments frequently employ certain atypical contexts or backgrounds inadvertently and in consequence will yield drastically divergent results; to an even greater extent this holds for the mediationally rather than functionally oriented, more traditional experiments that thrive on the infirming instances of a cue and programmatically employ grossly distortive configurations, such as most recently the situations created by Ames (1946).

Each of these experiments, may it be typical or atypical, is nevertheless indisputable in its results. But at the same time it is of unscrutinized ecological generalizability. The core situation of any systematic experiment or experimental setting constitutes what may be called a variate package, that is, a more or less incidental combination of specific values along a large and practically unknown number (x) of dimensions. Each such situation may be projected somewhere in the x-dimensional manifold of design which we have exemplified for two dimensions in figure 8. Each imaginary point—or small, orderly group of points—in such a space represents a potential systematic experiment.

Mostly there is little technical basis for telling whether a given experiment is an ecological normal, located in the midst of a crowd of natural instances, or whether it is at the fringes of reality, as are bearded ladies and other freakish creatures, or whether it perhaps is like a mere homunculus of the laboratory out in the blank. As a matter of principle, a study of one or of a series of individual sample situations—regardless of how life-like they may be—cannot answer a functional problem such as that of the degree of perceptual constancy. This holds true in spite of the fact that by the use of responder-replication or by systematic variation their results may become generalizable in certain directions and standardized for testing purposes.

No matter how much the results of systematic experiments may anticipate those of a representative survey, in a technical sense only representative design can answer functional problems definitively. As we have said, the ecological generalizability of the principle of perceptual compromise, or of "approximate" size constancy as originally suggested by Hering, has been established in this manner at least for one subject, along with the broader principle of distal rather than proximal focusing in the area of size with distance variant.

2. LOUDNESS CONSTANCY WITH DISTANCE VARIANT

Sometimes an entire array of individual systematic experiments may appear to be laid out after quasi-representative principles so as to cover the ecology or the mediational pathways by a vaguely conceived "one of a kind" rule. This tacit spot-checking or informal sampling procedure forms an ecological counterpart to what polling statisticians might recognize as a most rudimentary form of stratified, "quota" or proportionate sampling. It is usually of the highly erratic type sometimes labeled "accidental" in statistics. This primitive type of coverage of the ecology we have designated as "canvassing."

In the area of the perceptual constancies basic principles found in the area of size constancy, such as that of distal focusing and of compromise, may be re-examined for shape- or color-constancy. This is an example of canvassing that involves a spot sampling of distal variables within the same sense department. The same principles may also be investigated cross-departmentally, for example, by including the auditory or the tactile-kinesthetic domain.

Loudness constancy with distance variant is a problem analogous to size constancy. The distal stimulus variable is intensity of sound at the source; it is measured by placing a microphone next to the place from which the sound is emitted. The proximal stimulus variable, intensity of sound at the ear, is measured by a microphone next to the observer. The two variables are diacritically separated by placing the source at different distances from the observer. Two attitudes are possible, one toward loudness of sound as emitted at the source, and the other toward loudness of impact at the ear. They correspond to the naïve-realistic attitude toward bodily size (b) and to the analytic, photographic attitude (p), respectively, in the size-constancy experiment.

An experimental study of loudness constancy was undertaken by Mohrmann (1939). In a sound-recording room that created conditions somewhere between open-air and indoor hearing but probably closer to the former, two loudspeakers were placed at different distances from the observer. The distances used were 75 cm, 750 cm, and an intermediate distance, 237 cm, the latter being the geometric mean between the two extremes mentioned. All the three possible combinations of two distances were used; the loudspeakers exchanged position from time to time to avoid ill effects of familiarization. Types of sound were canvassed as to familiarity and/or complexity; five variants, ranging from speech over dance music, metronome beats and an explosive type of noise, to a relatively pure sine tone (c', 256 cycles) were employed. All sounds were piped into the two loudspeakers in serial alternation (speech and music from phonograph records), with successive presentation times of 2 seconds and pauses of ½ sec. between front and rear. (A schema of the electrical equipment is given in fig. 3, p. 164, of Mohrmann's paper.) The adjustment method was used, the subject turning a knob until the two sources—or else the two impacts—appeared to be equally loud (which took an average of 6.8 pairs of alternations). Experiments were performed as follows: in complete darkness prior to the subject's having any acquaintance with the room or with the arrangement of the speakers (first day), or blindfolded between the visual experiments or with a full visual view of the situation (second to fourth days). There were 24 adult subjects with normal hearing in both ears.

Figure 18 shows the results. Small circles represent mean constancy ratios for all three distance combinations taken together, horizontal lines, ranges of average points of subjective equality for the individual subjects. The natural attitude toward source seems generally more successful (over-all average $c = .82$) than the proximal or sensorial attitude toward impact at the ear (over-all average $c = .34$, which is farther from 0 than .82 is from 1.00). Thus the prevalence of distal focusing is confirmed.

Under favorable conditions, including the presence of visual cues for space and for the position of the sources, the average c-ratios are .78 or over and approach 1.00 in the case of sounds of high complexity and/or familiarity, notably speech ($c = .99$). Even when the non-visual series are included, speech constancy averages as high as .95. On the other hand, the average of the visual c-ratios in attitude toward ear is only about .50; this value drops to about .20—an improvement in view of the proximal aim—in the darkness series which support the reduction of space.

Results in the case of blindfolding are between those for the visual and for the darkness series. The two non-visual series would probably be closer to each other were it not for the presence of a practice effect which acted in favor of the blindfold series (see above) and which we will discuss in § XIII (fig. 30).

The three distance combinations yield very similar results for attitude toward source, but attitude toward ear becomes more efficient when the sources are relatively close (75:237 cm) and visual cues are lacking.

No effort was made to establish the exact nature of the cues by which distance was mediated under each of the various conditions. In line with the policies of the functional approach we did not probe into mediational technology beyond making sure of the presence of cross-departmental interaction as revealed by the significant improvement of constancy through the addition of visual orientation factors. It may perhaps further be relevant in this context that in special series, employing intensities barely above threshold, loudness constancy was improved, while it suffered when intensities were increased to an extent seriously interfering with the natural timbre and overtone relief.

In other special series the pause between alternating presentations was increased from ½ sec. up to 6 seconds, with the result that it became almost impossible to realize the attitude toward impact at the ear. This part of Mohrmann's experiment bears some indirect relation to the one by Joynson (1949) which we have mentioned earlier in this chapter and in which the role of the angle of separation as a support for size constancy has been ascertained.

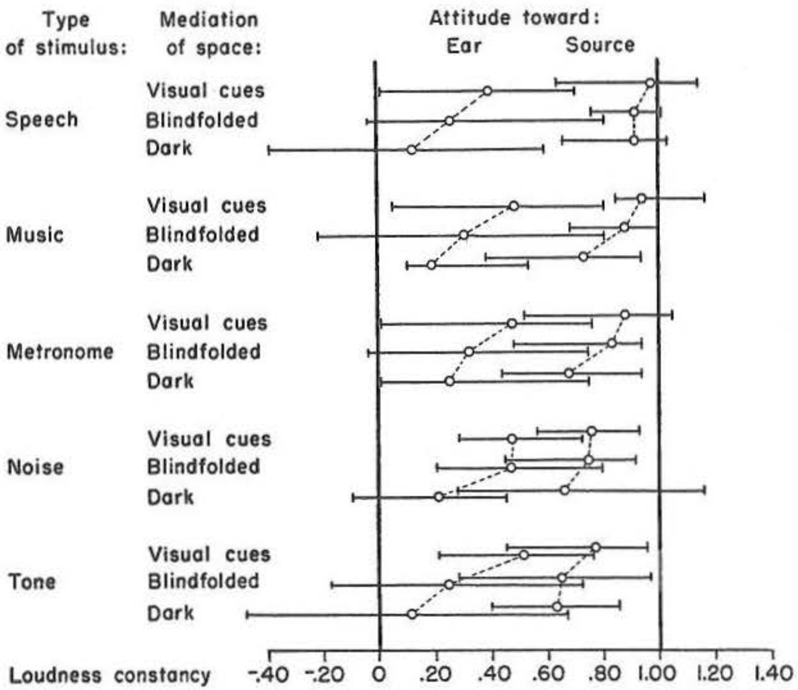


Fig. 18. (Redrawn after Mohrmann, 1939.)

Perceptual Loudness Constancy with Distance Variant.—In our redrawing the portions of the ranges outside of 0 and 1.00 are drawn to the same scale as those inside while they had been compressed to half-scale in the original publication.

3. A MULTIPOLAR CONSTANCY PROBLEM:
WEIGHT CONSTANCY WITH SPEED AND KINETIC ENERGY VARIANT

That there is a "world" of touch as there is one of vision has been made clear by David Katz in a pioneer monograph (1925) that parallels his more famous volume on the world of color (1911; 1930). In this tradition such problems as that of the apparent weight of submerged bodies or of the relative invariance of weight impressions with leverage of the arm variant (Fischel, 1926) were studied.

Another of the many conceivable tactile-kinesthetic constancy problems is posed when the intensity of the proximal impact is made variant by allowing

bodies to fall from varying heights and thus to impinge upon the receptors at differing speeds. As in the preceding examples, it is weight (or mass) that furnishes the distal variable by virtue of its landmark character in the identification of objects and its importance in the context of manipulation and orientation.

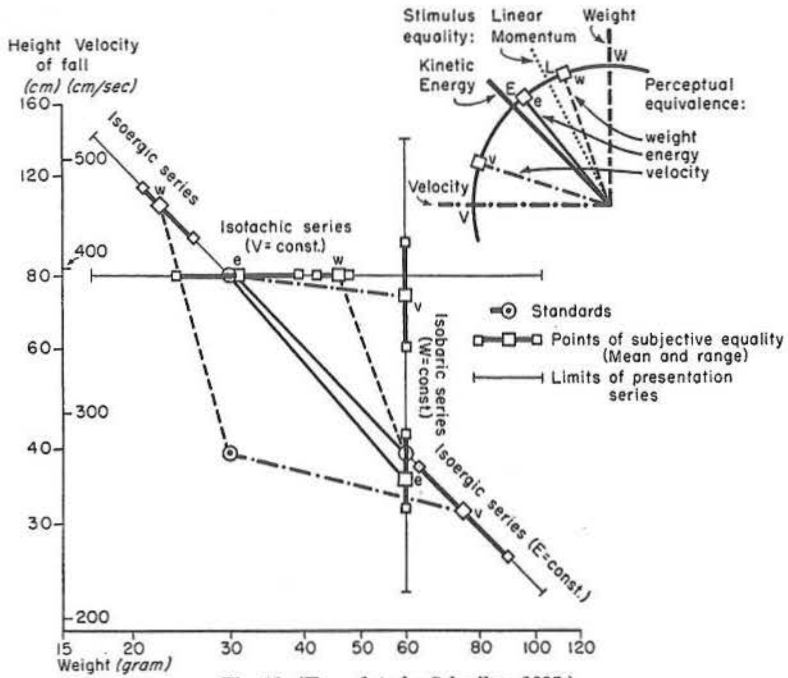


Fig. 19. (From data by Schreiber, 1935.)

Perceptual Attainment of Weight, Speed, and Kinetic Energy of Falling Bodies.—Presentation in each individual series started from the Comparison object representing the POE for the attitude called for in that series; thus possible influences of "adaptation level" (Helson) were working against rather than in support of the hypothesis of perceptual compromise which was to be tested.

This problem of weight constancy with impact variant was studied at Vienna by Livia Schreiber (see brief report in Brunswik, 1934, p. 161 f.). Balls of varying weights were allowed to impinge upon, and to roll off in rebound from, the prominence at the base of the thumb (thenar eminence) of the hand held at 45°. Weights and heights of fall (velocity) were arranged as indicated along the two main axes in figure 19. As in the experiment on loudness constancy, the mediation of height and speed was in some series confined to the intradepartmental—in this case the mechanical—cues, but figure 19 shows only results from a series in which visual and auditory cues were also present. Three perceptual attitudes were induced, toward weight (w), speed (v), and force of impact (kinetic energy; e). There were fifteen adult subjects. To simplify our presentation, only part of the Standards and

Comparison series are shown in figure 19. (Note vertical, horizontal and oblique directions of variation, called isobaric, isotachic, and isoergic.)

Results are indicated by the direction of the lines connecting the Standards with the mean PSE's on the respective comparison series, using broken lines for the two perceptual weight-equivalence contours, solid lines for the perceptual kinetic-energy contours, and dash-dotted lines for the perceptual speech contours. The ranges are also shown.

The abstract of the results at the top right of figure 19 shows best attainment for kinetic energy, least successful attainment for weight. But the fact that there are at least three rather than only two POE's or poles in this experiment, V, E, and W, of which the first and last are in a marginal position, may have something to do with this outcome. Two variables may "gang up" against the third in designs of this kind, which we will call "multipolar." Application of the constancy ratio is rather pointless in such cases. We will come back to a discussion of the complex interplay of factors in such multipolar systems in the next chapter.

Results are of the same general kind with visual cues absent, but the constructive interdepartmental influence of these cues is evident from a lessening in the degree of weight constancy when they are removed.

In interpreting the results, one must first consider the nature of the proximal impact. Visual cues as well as the initial dent on the palm during the first time differential are very nearly proportional to speed. In the later phases the deformation effect and the forced movement of the hand may become more nearly proportional to kinetic energy (or to the linear momentum, defined in physics as mv). Since the ball is not allowed to come to rest on the hand, weight is never directly represented. It thus remains untied to cue and stays a distal variable in our experiment. Even though weight is not attained as closely as is energy, each under its own attitude, any departure from proximal focusing in the direction of weight attainment must, as always in analogous cases, be seen as evidence of a trend toward distal focusing.

XII. The Extended Constancy Problem

The three traditional perceptual constancies (size, shape, color) as well as their analogues in other sense departments (loudness, weight) which were discussed in the preceding chapter have in common that the variant condition against which an invariance is upheld in approximation does not permanently adhere to the objects in question; rather, this variant condition is in the nature of a transient, extraneous relation or "circumstance," such as distance or tilt relative to the perceiver; illumination; or speed. An extension of the perceptual constancy problem is given by proceeding toward the problem of invariance relative to a certain bodily characteristic when another bodily characteristic in the same objects is allowed to vary

1. Weight vs. Density Constancy

The well-known size-weight illusion by Charpentier may be subsumed under this scheme. The problem here may be phrased as that of weight con-

stancy with volume variant, volume being a permanent (or quasi-permanent) property of the same objects whose weights and weight perceptions are under consideration. Contrast this with the case of Schreiber, just reported, in which such extraneous circumstances as height and thus speed of fall were the disturbing conditions in the perception of weight.

Subsuming the size-weight illusion under the concept of perceptual constancy raises the problem of a second pole of intention. In our case this is given by density (D), defined as weight per unit volume (or W/V). Two

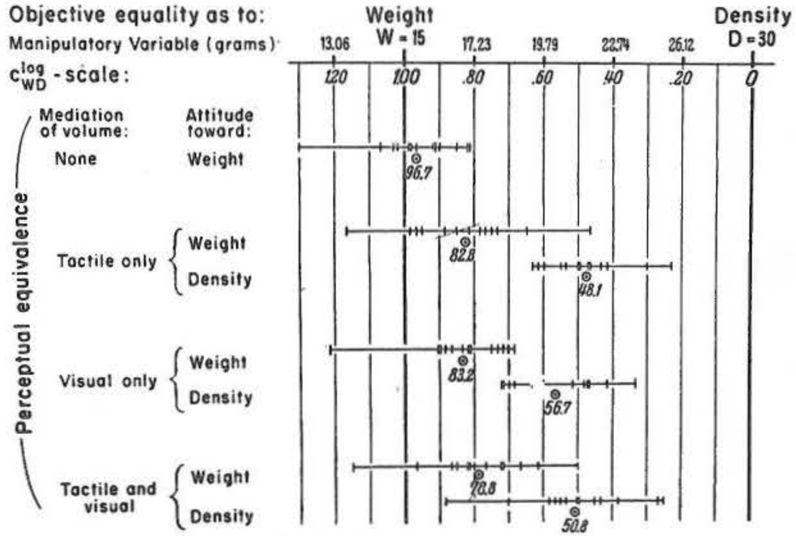


Fig. 20. (After Izzet, 1934.)
Perceptual Weight vs. Density Constancy with Volume Variant.

perceptual attitudes then are possible, one toward weight (w) and the other toward density (d). The former may be elicited by suggesting the fictitious case that the objects to be compared are to be weighed on a balance, the latter by asking which object would seem to float more readily on water.

An experiment by Izzet (1934) which was conducted under these auspices is summarized in figure 20. Matchboxes of equal shape but of a volume ratio of 1:2 were used. A small box weighing 15 grams was the Standard. The two poles or points of objective equality (POE) on the Comparison series are, then, $W = 15$ and $D = 30$. Standard and Comparison were placed simultaneously in the palms of the hands held outstretched from the elbow, right and left switching places in a random balanced order from series to series. A variant of the method of limits was used; the Comparison weights employed are listed at the top of figure 20. In some series the boxes were placed in the open hands of the blindfolded subject so that the mediation of volume was tactile only. In others the same experiment was performed in plain sight of the subject, leading to tactile-visual mediation of volume. And in still others the mediation of volume was restricted to visual cues by making the subject suspend the weights from the heads of pins attached to the boxes and placed between the fingertips, with the fingers pointing downward.

The center part of figure 20 shows individual PSE's averaged for each of the two attitudes and three types of volume mediation, for 13 subjects. Means are shown underneath each range and indicated numerically in terms of the constancy ratio, with *W* becoming 1.00 and *D* becoming 0. Weight constancy is higher (about .8) than density constancy (about .5) under our conditions. Since both weight and density are distal variables relevant to life, this is not at variance with our general contention of distal focusing. It also seems more appropriate to express results of constructive perceptual activity in the positive terms of constancy rather than in the more indistinct terms of "error" or "illusion." Note further that visual mediation of volume is about equivalent to tactual mediation or even to a combination of the two, again pointing to the intersensory character of the constancy mechanism.

The top row shows the results of a control series in which both of our major mediational avenues for volume were eliminated. The fact that weight attainment is only approximately perfect (.97 rather than 1.00) under these conditions points to the presence of some residual minimal cues for volume.

The drastic "over constancy" values at the extreme left in the rows with weight attitude represent the same individual in all four cases. This subject is the only one to reverse the direction of illusion from the customary weight-volume dissimilation to an assimilative effect.

2. VOLUME VS. SURFACE CONSTANCY

Another experiment in the category of the extended constancy problem concerns the perceptual volume- and surface-constancy of three-dimensional bodies with shape or the proportions of height and width variant. This experiment was undertaken by T. I. Stevenson (for a summary see Brunswik, 1933, and 1934, pp. 152–155). A sphere and four square prisms were the Standards used with a series of cubes of varying size in simultaneous comparison (fig. 21). The volume of each of the Standards was equal to that of a 7 cm Comparison cube; the heights of the prisms ranged from 70 cm ("long stick") to .7 cm ("thin plate"), and their base-squares from 2.2 to 22.1 cm (see the scale in fig. 22).

Results, presented in figure 22, show that the mean PSE's as well as the ranges (12 adult subjects) for volume- vs. surface-attitudes are very close to each other for elongated prisms, while there is considerable attitudinal differentiation (shift span) for flat objects. Since the relative position of the two intended POE's (*V* and *O*) is similar for the elongated and for the corresponding flat bodies, a synoptic interpretation of all results is possible only on a more than two-polar basis. We have chosen base and height for additional consideration so as to be able to describe the results in terms of perceptual compromise, but since here several variables are interdependent in a strictly mathematical manner, the choice of poles and indeed the number of independent factors are open to discussion. As we have said, the constancy ratio can not meaningfully be applied in such complex cases of multivariate interaction.

Note further the small variability (low threshold) and almost vanishing "constant error" for cube-with-cube comparison (center). The reader will recognize this as the case of the classical psychophysical experiment (§ III). It is a case drastically reduced in its research potential, as we have seen,

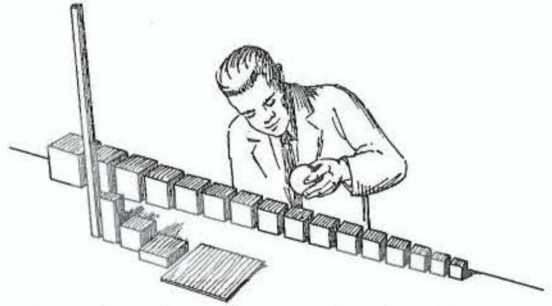


Fig. 21. (From data by Stevenson, after Brunswik, 1934.)
Objects Used to Test Volume and Surface Constancy.

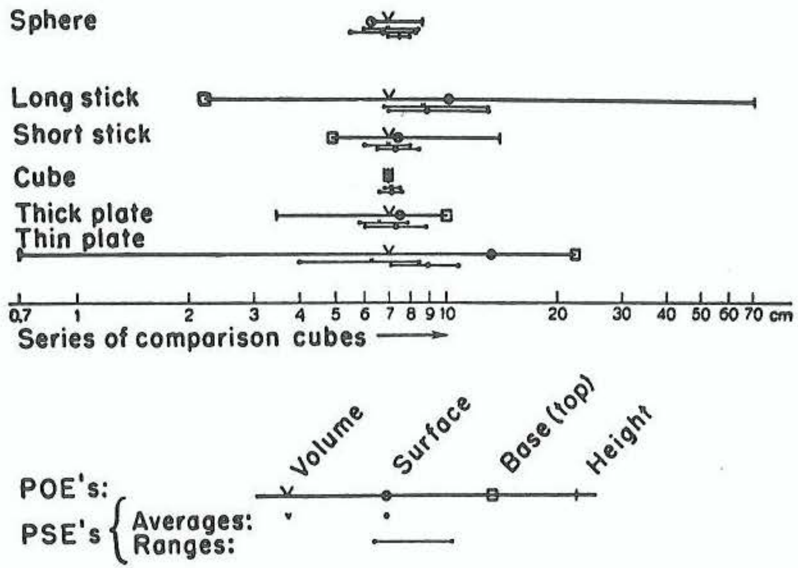


Fig. 22. (From data by Stevenson, after Brunswik, 1933, 1934.)
Volume and Surface Constancy with Shape Variant.

confining as it does the perception problem to the threshold problem under the most safeguarded, near-measurement conditions and discouraging the constant error which in turn is the most promising starting point for higher-complexity work in perception.

Some important implications of the Stevenson experiment with respect to the theory of distal focusing will be taken up in the next chapter (§ XIII/3).

3. PERCEPTUAL VALUE CONSTANCY AND MONETARY VALUE AS A DISTURBER OF NUMEROSITY PERCEPTION

A third example of an extended constancy problem is that of value- vs. numerosity- vs. area-constancy. Like the preceding example it is multipolar and thus defies the application of the constancy ratio. Its outstanding characteristic is that monetary value has been made one of the distal variables (see p. 16). The fact that values may be classified with the emotional and motivational factors, and that it may be interesting to know whether such factors can act as disturbers of the cognitive process, was but one—and to us the less commanding—of the reasons to study their place in perception. That such disturbances do occur has been known for some time. The macropsia and micropsia of emotionally loaded objects was described both by psychopathologists and by E. R. Jaensch; and reportedly the enlargement in the drawn size of bills of the larger denominations had been the subject of one of the demonstrations in Woodworth's laboratory course over a number of years. The well-known Bruner and Goodman experiment (1947) has followed up this problem, and Bruner and Postman have stimulated a long series of research along this line, recently surveyed by Postman (1953).

To those of us who see in functionalism mainly an avenue to the study of veridicality in perception the emphasis is shifted from the negative aspect of value-induced error to the positive aspect of value constancy. From this latter point of view the monetary value of such objects as coins or stamps is a distal variable worth knowing in its own right. The intriguing part of the problem is that monetary value exists by *fiat* of the "laws" of a cultural ecology only; thus it is of a different, more limited and more temporary character than are purely physical object characteristics and the natural or geometrical laws by which they are interconnected. This in turn means that value-cues must be acquired cues. In consequence, the study of value constancy as well as that of the interference of values with the perception of other object properties may shed light on the role of learning or familiarity in intuitive perception.

Of the two relatively early experimental studies undertaken under the direction of this writer in the area of value perception one (by Zuk-Kardos; see Brunswik, 1934, pp. 140–149) has used Austrian stamps with Viennese subjects, and the other (by Fazil; see Brunswik, 1934, pp. 147–150) Turkish coins with Ankara subjects. We will confine our attention to the latter. Figure 23 shows at the top the Standard group of 40 coins corresponding to a would-be American "half-nickel." At the bottom are shown the three POE's composed of what corresponds to the American "quarter;" $N = 40$ is objectively correct as to numerosity, $A = 16$ is correct as to sum of areas covered by the individual coins, and $V = 4$ is correct as to value. There were 22 adult subjects thoroughly accustomed to using these coins; relatively brief exposure of stimulus aggregates in simultaneous comparison precluded explicit counting.

Results are shown in the center row of figure 23 and in figure 24. Value

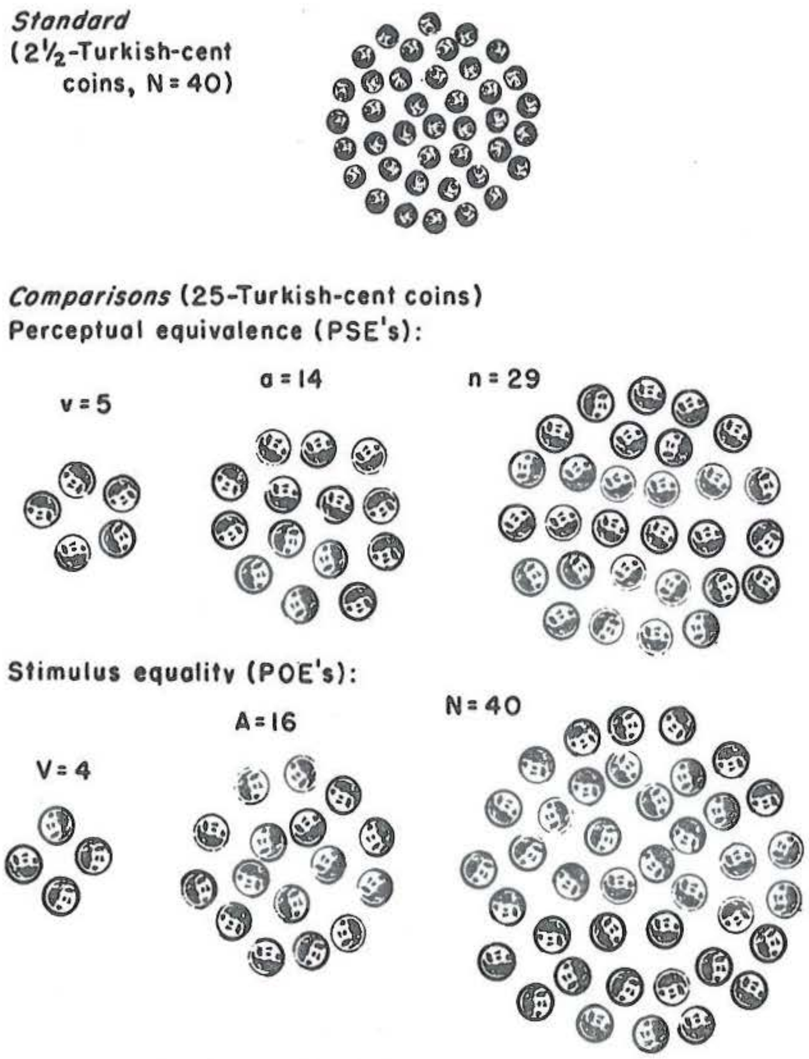


Fig. 23. (From data by Fazil, after Brunswik, 1934.)
Perceptual Attainment of Number vs. Area vs. Value.

constancy is considerable, but there is some compromise with area and/or number. In turn, the latter variables are affected by value; this becomes clear by the position of a to the left of A rather than toward N in figure 24, as well as by the results of some preliminary experiments using plain disks instead of coins in which both a and n were farther to the right.

Overestimation of the size of more valuable coins, rather than of their numerosity in groups as shown here, was also ascertained by Fazil, but differences between rich and poor subjects were not studied prior to the study of Bruner and Goodman (1947).

Zuk-Kardos had found that a newcomer who "knew" the value of Austrian stamps but had not yet developed a second-nature type of handling familiarity with them did not show value constancy or the corresponding illusion of number. By using both American and Canadian stamps with both American and Canadian subjects, Ansbacher (1937) was able to confirm the crucial part of our results as well as further ascertain the role of familiarity. The autonomy of perception and of perceptual learning will be discussed in greater detail in § XVII.

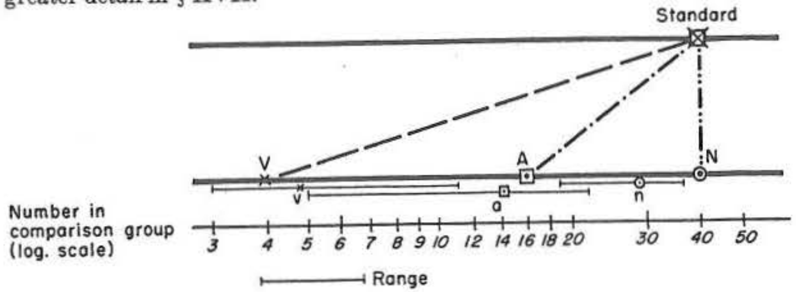


Fig. 24. (From data by Fazil, after Brunswik, 1934.)
Three-polar Presentation of Number-, Area- and Value-Constancy as in Fig. 23, Adding Ranges.

4. THE PERCEPTUAL SYSTEM AS AN INTUITIVE STATISTICIAN: THE PERCEPTION OF VARIABILITY

At first glance the constancy achievements hitherto presented appear like simple computations involving but a few constituent variables and univocal relations between them. Yet the fact that systems of alternate cues possessing limited ecological validity may be involved in the "registering" of such circumstances as distance or illumination injects a statistical element into these implicit computations. The performances in social perception reported in § VI are strongly suggestive of statistical integration processes. Attneave (1954), and Hochberg and McAlister (1953) have recently made much of the statistical nature of form perception. This may involve the ability of the organism to utilize simple lawful relations that hold within certain groups of stimuli in the visual field (say, within part of the contour of an ink bottle)—"local laws" as this writer (1934) has called them. Or it may involve the organism's ability to impose such relations upon an incomplete or recalcitrant stimulus collective. Ivo Kohler's findings with distortive spectacles (1951) benefit from interpretations which compare the "frequent" and the "rare" in our surroundings. It may even be that perceptual compromise is but one of the manifestations of the uncertainty of the ecology and of the resultant basically statistical nature of perceptual functioning.

All this presupposes that the perceptual system be capable of tacitly evaluating the statistical properties of groups. It is therefore fitting to conclude this chapter with a brief report of a study by Hofstätter (1939) in which the absorptive and representational power of the perceptual system relative to sets of data and to concepts usually considered the exclusive domain of the statistician was made the major topic. The execution was under the auspices of the constancy problem. Groups of randomly arranged sticks of equal thickness but varying length, held together in bundles by a rubber band, were used in all the experiments. Results based on about 30 subjects revealed that not only can such higher-order variables—as Gibson would call them—as the mean or the standard deviation be appraised intuitively, but that perception spontaneously progresses from there to an intuitive appraisal of the coefficient of variation (V = S.D./Mean as defined by Pearson).

In one of Hofstätter's series a Standard bundle of five sticks (4, 6, 8, 10, and 12 cm long) was compared with a Comparison series of nine bundles, No. 2 of which (20, 18, 16, 14, 12 cm) represented the POE as to standard deviation, and No. 7 of which (24, 20, 16, 12, 8 cm) represented the POE as to coefficient of variation. The average PSE was at No. 6.3. This yields a "perceptual V-constancy with mean length variant" of .86 (in terms of the logarithmic constancy ratio). As can be seen from the bottom curve in figure 25, the compromise of V-constancy with S.D.-constancy is in this case revealed by skewness only, while the mode coincides with V-constancy.

Hofstätter proceeded to a systematic untying of a proportionality factor which was a surreptitious aid to V-constancy by substituting other Standards with the same S.D. but different shape of distribution (one S′, composed of 4.7, 5.7, 7.3, 9.8, 12.5 cm, and the other, S″, composed of 4.5, 5.0, 8.5, 10.5, 11.5 cm). This diacritical design led to partial reduction of V-constancy to the less complex S.D.-constancy; there is a marked second mode for S″ at S.D. = const., and the perceptual S.D.-constancy with mean length variant reaches .56.

Even eight- to ten-year-old children perform well on such tasks (provided that a translation of the instruction into suitable terms of the child's daily experience can be found). Furthermore, statisticians with years of experience succumb to the same illusions as the other subjects. Still further, there are influences from irrelevant factors (such as thickness), and the correlation between intuitive performance and intelligence tests measuring related abilities is low. All this supports the view that the performances are essentially perceptual in character rather than being in the nature of a stimulus error based on an intellectual approach.

In particular, it was possible to ascertain that the perceptually just noticeable difference for standard deviations lies somewhere in the neighborhood of .08. Furthermore, the exponent applied to the deviations in the implicit perceptual computations could be reconstructed from the results; usually it is not 2, as in the customary mathematical formula, but tends to be lower (1 or even ½) if there is a good deal of variation in the middle range and higher (up to 6) where variation is concentrated at the extremes.

In another publication, Hofstätter (1940) has applied some of his ideas to a study of the formation of the reputation of the intelligence level of regional groups, using recruits from the various provinces of Austria. He found an interpersonal mechanism of statistical information-gathering at work in which personal contact plays a minor role. In view of his finding that emotional factors tend to exert a distortive influence (such as in the overestimation of one's home province), the conclusion that reputation may be used as a substitute for direct fact-gathering seems to overshoot the goal, however, at least where social responsibilities are involved.

The present chapter has assembled evidence of the impressive versatility of the perceptual system in seeking out the relevant factors or compounds in the environment however remote they may be or however complexly they may be constituted logico-mathematically. At least this is what seems to follow from a casual canvassing of the possibilities concerned.

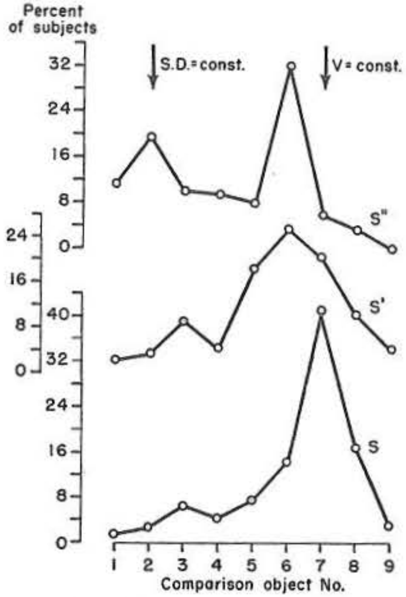


Fig. 25. (After Hofstätter, 1939.)
Intuitive Estimates of Variability.

XIII. Developmental Aspects of the Constancy Problem

In this chapter we will take up developmental problems in the field of the constancies. From ontogenesis we will proceed to the artificially telescoped developmental sequences furnished by practice experiments; from there we will move on to suggesting an over-all scheme of perceptual differentiation. This scheme will be used, in the subsequent chapter, for a functional differentiation between perception and thinking. In all this we will cut across from one distal variable (type of constancy) to another and from one sense department to another in the manner we have become accustomed to in the preceding two chapters.

1. Ontogenesis of the Perceptual Constancies and Its Ecological Generality

As we have seen in discussing intercorrelations and reliabilities in the area of the perceptual constancies (p. 22 f.), thing constancy does not seem to be a highly unitary function in the sense of differential psychology. This does not imply, however, that group trends should not exhibit greater uni-

formity when ecological comparisons involving different kinds of focal variables are the issue. Figure 26 shows the results of three studies in the development of the constancies. The constancy ratio is used to make the results more comparable. The first of the studies is by Beyrl (1926) and deals

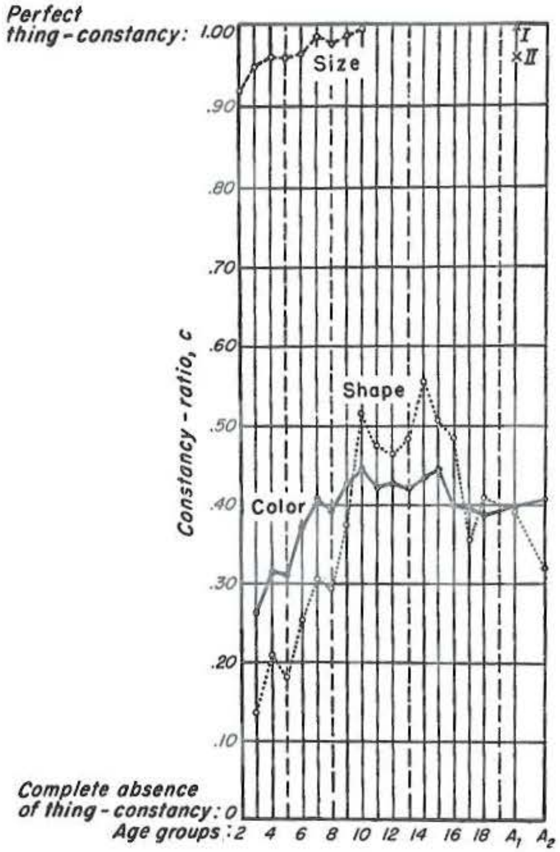


Fig. 26. (From data by Beyrl, Brunswik, and Klimpfinger; after Brunswik, 1930.)
Similarity of Development of Three Traditional Constancies.

with size constancy; disks and cubes were used at distances ranging from 2 to 11 m. Another study is by this writer (1928) and deals with whiteness constancy by using a series of gray papers ranging from white to black set up at varying distances from a light source ("illumination perspective"). The third is by Klimpfinger (1933*b*) and deals with shape constancy. She had used an ellipse (20 cm high and 30 cm wide) rotated from the frontal plane by 33½° about its vertical diameter as a Standard and a series of frontal ellipses as Comparisons; the *B* pole was at 30 cm and the *P* pole at 25 cm horizontal diameter.

The age ranges in figure 26 are from 2 or 3 years up to younger and older groups of adults (A_1 and A_2). Beyrl's data are limited to ages 2 to 10 and a group of rural adults (I). Since Beyrl's original sample of adult subjects was too small for significance, Klimpfinger (1933*b*, p. 624) later tested 20 additional adults under somewhat different conditions (II); this value, slightly misplaced in this writer's original graph (1930), is corrected here. The dotted vertical lines in the figure refer to the boundaries of certain developmental phases distinguished by Charlotte Bühler (1931) which shall not further be discussed here.

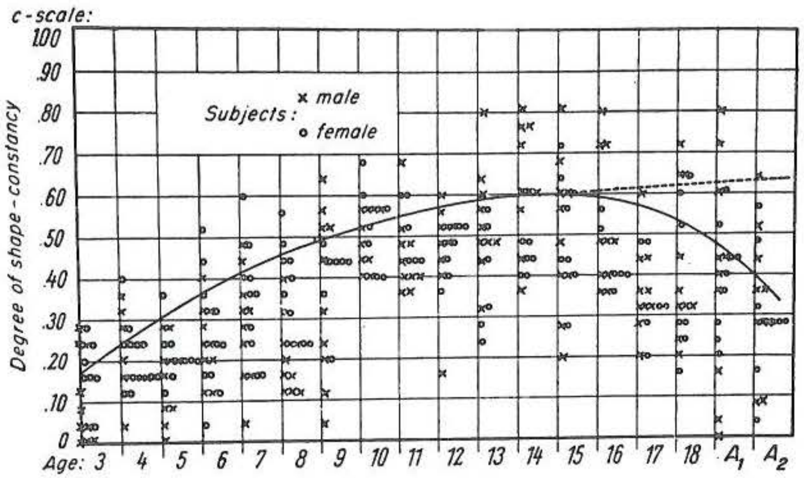


Fig. 27. (After Klimpfinger, 1933*b*.)
Individual Differences in the Development of Perceptual Shape Constancy.—
Curves are freehand fits to over-all trend in 313 subjects, age 3 to 70.

The fact we wish to repeat here from our preliminary discussion on page 22 is the parallelism in both the upward swing and the regression of the curves. Both rise and fall are statistically significant for color ($n = 232$) and shape ($n = 313$; for breakdown by individuals see fig. 27). As suggested by the freehand broken line in figure 27, a certain subgroup of adults (see below) may not participate in the decline.

Note further the differences in absolute levels of achievement for the three kinds of constancy. This is occasioned in part by a "systematic" effort to obtain maximum interindividual differentiation by inducing conditions that will lead to intermediate levels of perfection. It is very likely, however, that the differences are in part "representative" of intrinsic differences in the difficulty of the functions concerned; obviously, size constancy benefits from gross depth cues while shape and color constancy must to a large extent fall back upon more subtle articulations.

The policy of deliberately setting the over-all perfection of performance at a level lower than the optimum so that factors playing upon that level may be brought out more dramatically has also been followed in Holaday's experiment on size constancy (§§V, XI).

This policy must be defended in view of the fact that some developmental studies which had shown evenly perfect constancy performance, notably that on color constancy by Burzlaff (1931), were interpreted as indicating that constancy as a whole was perfect throughout (see this writer's Note in Klimpfinger, 1933*b*, pp. 619–622; see also Koffka, 1935).

For methodological augmentation we briefly summarize the results of a recent study on size constancy by Dukes (1951) in which representative design was for the first time applied to a child, a 6-year-old boy. Figure 28 presents both design and results in a

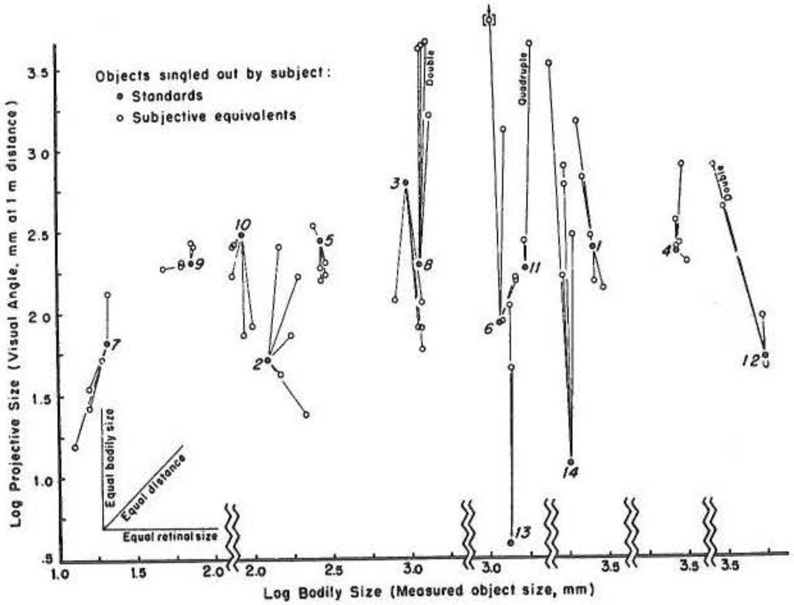


Fig. 28. (Rearranged from Dukes, 1951.)
Representative Study of Perceptual Size Constancy in a 6-Year-Old Boy.

rearrangement that facilitates comparison with our figure 8. Like our adult subject, Dukes' boy was randomly interrupted in his daily activity and asked to designate the object he was looking at; this was done 14 times, yielding an equal number of "Standards." Instead of a numerical judgment, the subject was asked in each case to designate up to 5 size-equivalents freely chosen from the field. This yielded an N of 67, involving all manner of objects, including houses and trees. Note from figure 28 that the subject chose size-equivalents from distances often greatly at variance with the distance of the Standard; in other words, he did not prefer the greater alleged "ease" of staying within the same distance for each act of comparison as would have to be expected in the tradition of classical psychophysics.

Figure 28 reveals the preponderance of vertical over horizontal error components, indicating that perceptual equivalences are fairly accurate as to bodily size while reflecting but little the equivalences as to size of retinal image. The correlation of the responses with bodily size is .991 (compare with similar values in our fig. 9), but with retinal size it is close to zero (.09).

2. TELESCOPED DEVELOPMENT AND DIFFERENTIATION OF THE CONSTANCIES THROUGH PRACTICE

Another experiment by Klimpfinger (1933*a*), using the same shape-constancy setting with tilted and frontal ellipses as in her developmental study just reported, undertook to investigate the effects of attitude and practice. Five adult subjects went through thirteen sessions spread over the period of a month. There were two series in attitude *b* in each session, and two additional series in attitude *p* in five of the sessions. The practice was unrewarded in that no information as to the correctness of the responses was given out.

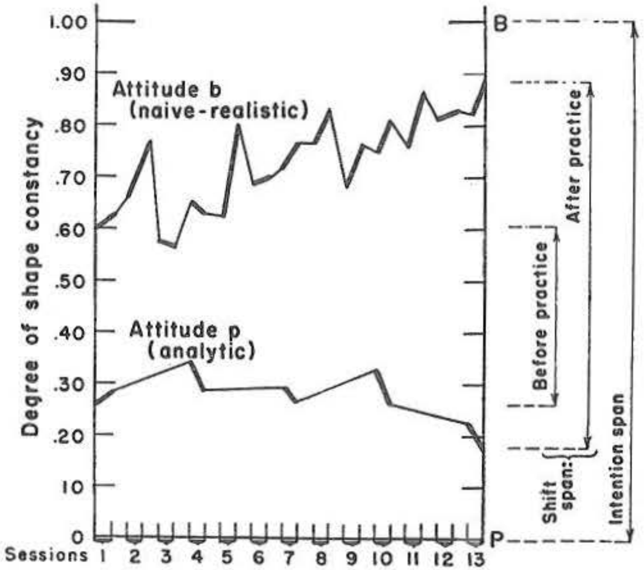


Fig. 29. (After Klimpfinger, 1933*a*.)
Influence of Practice upon Perceptual Shape Constancy.

Figure 29 shows the results. They give the picture of a telescoped type of development. To the summary given on page 22 we may add that the eventual downward swing of the *p*-curve is preceded by a slight upward trend. This trend, following as it does the direction of the *b*-curve, may be interpreted as a lack of differentiation in the earlier stages of our practice development. In the later stages *p* becomes more and more differentiated from *b*, which means that both attitudes show improvement relative to their own respective poles of intention.

For comparison we present come additional data from the study by Mohrmann on loudness constancy reported in § XI/2. A survey of results of some series that were repeated on three successive days (fig. 30) shows a trend similar to the first part of the Klimpfinger graph; her over-all results thus do not seem to be ecologically fully

generalizable but contingent on content and/or the position of the respective function in the developmental hierarchy.

In both figures 29 and 30, repetitions of the same series within one session are joined by heavier lines within the curve. We notice a rather consistent trend toward exaggerated progress along the main direction of development for the respective attitude at the given period of time. With few exceptions the heavy portions of the curves rise rather markedly where the over-all trend is upward and drop markedly where the over-all trend is downward. This fact brings out a recency effect in the telescoped practice developments of the constancies.

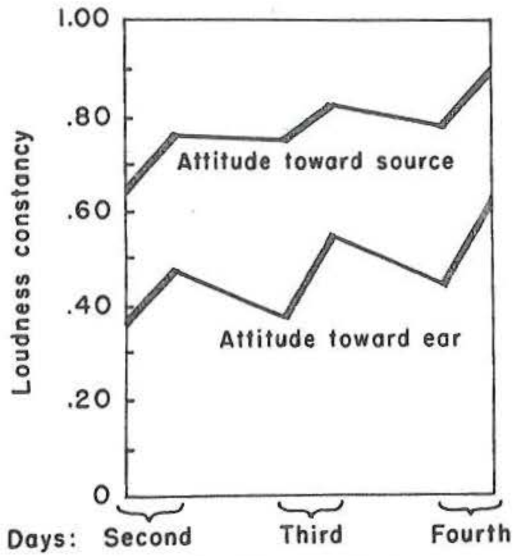


Fig. 30. (After Mohrmann, 1939.)
Progress in Loudness Constancy.

3. A GENERAL DEVELOPMENTAL SCHEME OF PROXIMAL VS. DISTAL RELATEDNESS

An ecological integration of the developmentally relevant facts presented in the present and in the preceding two chapters is attempted in figure 31.

Since distal focusing requires a complex instrumentarium, we must assume undisturbed proximal relatedness at the beginning. This may be illustrated by Uexküll's example of the sea urchin (fig. 11), by von Senden's (1932) data on the highly defective space perceptions in cases of operated congenital blindness (see also Hebb, 1949), or by the writer's and Cruikshank's (1937; for full report see Cruikshank, 1941) finding of a complete absence of size constancy in children at about four months of age.

As we have said (p. 22), there are some inklings of a beginning development toward size constancy at about six months. But it must be assumed that attitudes are not very well differentiated at early stages of each of the various constancies. In adults Stevenson was able to capture volume- vs. surface-constancy in such an early stage in connection with some cases of drastic shape discrepancy (cubes vs. sticks; fig. 22). The fact that constancy is meager and variability high under such circumstances, and that in addition subjects find the task difficult and unpleasant to the point of embarrassment, bespeaks the fact that the perceptual system indeed functions at a lower level in such tasks.

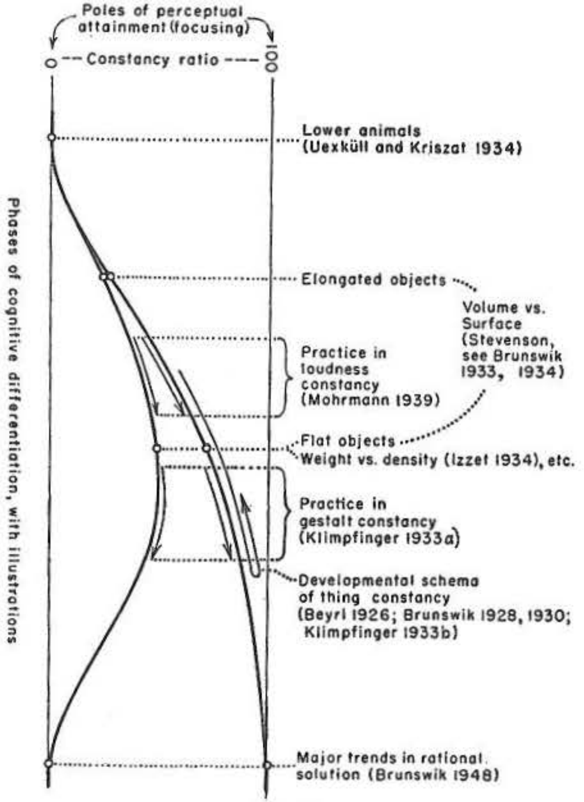


Fig. 31
Schema of Cognitive Differentiation, with Examples.

Limited differentiation with a great deal of compromise is given by Mohrmann's results, discussed in the preceding subsection, by Stevenson's results with flat objects, and by the first half of Klimpfinger's practice experiment.

The second half of Klimpfinger's experiment marks the beginning of fuller differentiation, one that in addition may be cultivated by differential practice.

The group of ontogenetic curves with their eventual regression is fitted to the upper of the two differentiating branches.

Toward the right-hand end of figure 31 the two theoretical developmental curves are joined with the respective poles of intention, the distal at the

top and the proximal at the bottom. This is an idealization, at least so far as distal perception is concerned. As we have seen, even in the most advanced forms of perceptual constancy some compromise remains; we have undertaken our representative design in size constancy (1944; summarized above, pp. 43–48) to find out whether or not this compromise tendency really holds, and came out in the affirmative. But as we will see in the next chapter, our idealization may be of greater cogency when applied to the case of the higher-order analogue of perception among the cognitive functions, that is, to thinking.

XIV. Perception and Thinking

In this chapter we take temporary leave of questions of representative design and of its various approximations in order to follow up some of the implications of our developmental scheme with respect to the general problem of the interrelationships between perception and the intellect proper. In this context we will also turn to the classification of perceptual attitudes and their relation to the so-called stimulus error.

1. Compromise vs. pointed distribution of response

Perfect attainment of the distal pole in figure 31 is possible only if the laws of physics are being applied to all the situations in question. The particular law involved in the case of size constancy has been stated at the center of the left-hand side of figure 10 (p. 51). It is easily seen that it can be applied only if a series of relevant conditions in the external medium is known. Only then can the confirming cases of a cue be sifted from the infirming cases, and a single cue—say, for distance—will suffice where perception needs many. A foolproof ascertainment of distal variables thus requires not only knowledge of the law but also manipulation and locomotion; there is no time for such instrumentality in "instantaneous" perception, let alone the availability of the law. Only the measuring and computing physicist can be reasonably "certainty-geared;" intuitive perception must remain "uncertainty-geared." Since perception is not equipped with the necessary added information, its performance must depend on relatively superficial and stereotyped cues of limited ecological validity, preferably a multitude of them; attainment can never be ideal under such circumstances. Even if on the average attainment of the pole should be perfect, considerable variability would remain as a result of the intrinsically limited validity of perceptual cues for distance or bodily size.

The distinction between certainty-geared and uncertainty-geared functioning is one by definition. But its concrete use in distinguishing between thinking and perception should be illustrated at least by example. We have therefore constructed a task in size constancy in two corresponding versions, one involving a typical case of perception and the other a typical case of arithmetic reasoning (for brief reports see Brunswik, 1948, 1954). In both cases 8 cm was the distally correct response (B) and 16 cm was the objective proximal equivalent (P). In the perceptual version a stimulus situation was

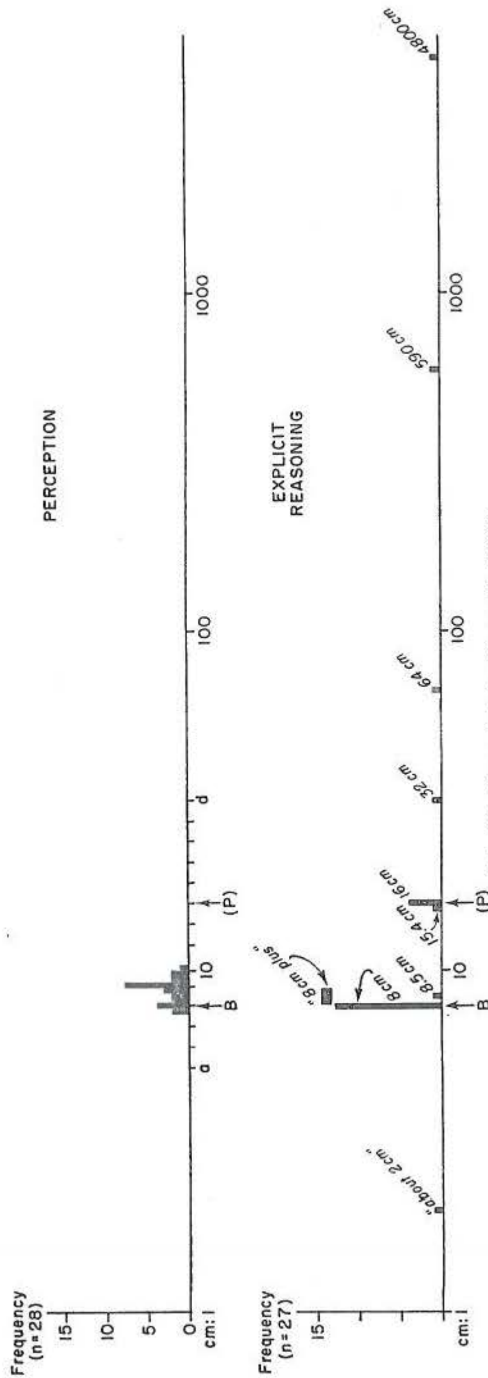


Fig. 32. (From data by Brunswik, 1948.) *Distributions of Accuracy and Error in Perception vs. Thinking.*

actually presented in the laboratory, with the normal array of distance cues left intact. In the intellectual or "thinking" version the question was made part of a subsequent examination, giving such numerical indications about the situation that the correct answer could be univocally ascertained. Partially overlapping groups of undergraduates (n = 28 and 27, respectively) served as subjects. All had been previously given an explanation of the problem and had been informed of the major trend of results of experiments in perceptual size constancy.

Results of the perceptual version of the task (fig. 32, top) show the usual trend toward "approximate size constancy" (*B*-pole) tainted with the characteristic slight compromise tendency toward the opposite type of task, photographic correctness (*P*-pole). All responses (PSE's) fall in a compact and fairly normal frequency distribution, with the geometric mean at 8.95 cm. This corresponds to a logarithmic constancy ratio of .84. It is to be especially noted that, as is usually the case under ordinary conditions of a thing-constancy experiment, there is no particularly outstanding frequency of precisely correct answers.

For the reasoning approach (fig. 32, bottom), on the other hand, almost half of all the answers (13 out of 27) coincide with *B* and thus are on-the-dot correct. A second discrete bar—of more moderate height, to be sure—represents on-the-dot correctness relative to a second major potential task, one not actually asked of the subjects; this is the finding of the retinal equivalent (*P*). It demonstrates one of the typical pitfalls of reasoning, namely, the going off in the wrong direction by being right about something else.

It may thus be said that while, in our case at least, perception lingers in the twilight zone of compromise, thinking shows some of the pointedness alluded to at the right end of figure 31. The intellectual approach, using measurement and calculation, thus appears to fulfill the ultimate ends of perception in a way perception itself is incapable of doing. In humans the two levels of cognition coexist, mostly in peace, sometimes in conflict. As we have mentioned on page 22, Werner has spoken of perception and thinking as "analogous functions" serving the same cognitive purpose at different developmental levels. The regression in the perceptual constancy curve after adolescence which we have discussed in the preceding chapter may be accounted for by the fact that in adults the needs for supreme accuracy are best taken care of by measurement and calculation, and that a modest drop in intuitive perception may thus be well afforded.

All this does not mean, as we have seen, that thinking is without error. The second bar, at (*P*), represents error in terms of the task actually asked for. In fact, for all the answers given in the reasoning version of the task, the mean PSE (logarithmic) is 14.7 cm, which corresponds to a c-ratio of only .12; and the S.D. is more than ten times that obtained for the perceptual version.

The entire pattern of the reasoning solutions in figure 32 resembles the switching of trains at a multiple junction, with each of the possible courses being well organized and of machine-like precision yet leading to drastically different destinations only one of which is acceptable in the light of the cognitive goal. This pattern is illustrative of the dangers inherent in explicit logical operations. The relative ease of switching off at any one of a series of choice points in a basically linear, unidimensional, all-or-none series of relays is at least in part the result of the precise formulation, yet *relatively* small number of basic cues, involved in most typical reasoning tasks. The combination of channeled mediation, on the one hand, with precision or else grotesquely scattered error in the results, on the other, may well be symptomatic of what appears to be the pure case of explicit intellectual fact-finding.

On the other hand, as we have seen, intuitive perception must simultaneously integrate many different avenues of approach, or cues. Whereas distance is directly stated in the reasoning version, its perceptual "registering" must remain based on insufficient evidence, that is, on criteria none of which is foolproof or fully "ecologically valid." This of course makes for the task not being exactly the "same" in the two versions, but merely "analogous." It is the insufficiency of single cues which must be seen as responsible for the establishment in perception of "cue-family-hierarchies" (the phrase coined in analogy to Hull's "habit-family-hierarchy," see Brunswik, 1952, p. 19). For depth, such a series of cues is shown in figure 10. All of them must be seen as contributing to the result, often in competition with each other.

A further feature, often noted by introspectionists in search of a distinction between intuitive perception and thinking, is the flash-like speed of perceptual responses. It is a biologically very valuable feature, especially where life is constantly threatened by sudden danger or where chances of success depend on quick action. The almost instantaneous promptness of perception could hardly be achieved without the stereotypy and superficiality in the utilization of cues which we have noted and which makes for a certain intrinsic "stupidity" of the perceptual apparatus (see Brunswik, 1934, pp. 119 f., 128, 223 ff.).

The various rivalries and compromises that characterize these dynamics of check and balance in perception must be seen as chiefly responsible for the above noted relative infrequency of precision. On the other hand, the organic multiplicity of factors entering the process constitutes an effective safeguard against drastic error. Even with extensive statistical sampling of size-distance combinations under life-like conditions as we have undertaken it in § VII/3, the over- or underestimation as to bodily size was no more than three- or fourfold in the most extreme cases (see above, p. 47) and thus is ordinarily within the limits of ½ power of 10 when translated into the logarithmic terms of our figure; this contrasts with an order of magnitude of three powers of 10 for our extreme example of a reasoning error. Ordinary perceptual illusions are known seldom to exceed average values of forty per cent, and drastic fooling of perception is rare except under highly artificial conditions. As we must hold against recent criticism by Ittelson (1951), the "stupidity" of perception thus is by no means to be construed to mean maladaptiveness; as we all know, life has survived on

relative stupidity from time immemorial, and if threatened in its existence it is so by malfunctioning of the intellect rather than by malfunctioning of perception.

Considering all the pros and cons of achievement, the balance sheet of perception versus thinking may thus seem seriously upset against thinking, unquestioned favorite of a culture of rational enlightenment as the latter has been. From the point of view of strategy, perception would likewise appear to have gained in stature by our realization of its inherent "vicarious functioning." So long as we accept, with Hunter (1928) and Tolman (1932), vicariousness as the foremost objective criterion of behavioral purposiveness, perception must appear as the more truly behavior-like function when compared with deductive reasoning with its machine-like, precariously one-tracked, tight-rope modes of procedure. The constantly looming catastrophes of the intellect would be found more often to develop into catastrophes of action were it not for the mellowing effect of the darker, more feeling-like and thus more dramatically convincing primordial layers of cognitive adjustment.

Ending on a note of caution, we should like to stress that the representativeness of our two versions of a common cognitive task is open to some doubt. Many specific conditions could be listed under which it is perception which is bizarre while it is thinking which is mellow and given to compromise. Aside from deductive considerations, only representative design could definitely prove us right or wrong in our conjecture that the juxtaposition which we have presented is more typical than its reverse.

As mostly in this book, however, our major aim is not content but the demonstration of methodological possibilities. In the present case this is the possibility of raising the distinction between psychological functions from the traditional introspectionistic to a more objective level by the use of a comparative statistical analysis of the strategy and of the achievement of cognitive functions in terms of error distribution.

2. STATISTICAL SEPARATION OF ATTITUDES

The separation of perception and thinking is but one of a number of disjunctive steps that may be taken in the functional classification and grouping of cognitive processes. A more subtle type of analysis concerns the separation of attitudes within the specific framework of perception. We shall now attempt such a separation in basically the same manner in which we have just attempted to separate perception from thinking, that is, by considering the distribution of response or error.

As will be remembered, Klimpfinger's (1933*a*) experiment on practice in shape constancy, presented in § XIII/2, involved two different modes of field structuring within the perceptual system proper. One was centered about "distal" bodily form, the other about the form of the photographic projections of the bodies concerned. In figure 33 the frequency distributions of the two respective sets of judgments or PSE's are shown in the same manner as in the practice curves (fig. 29). Since the analytic, sensorial attitude, i.e., the attempt to suppress the natural mechanism of shape constancy, was induced less frequently than the natural attitude, the corresponding frequencies have been multiplied by a constant factor to obtain comparability.

If we forget for a moment that different instructions were given to elicit the two attitudes, and also renounce efforts towards an introspective ascertainment of attitude taken, there is nothing left but the composite curve enveloping the two component curves. Its bimodality, which is clear-cut in spite of the leveling effect of the systematic improvement that takes place

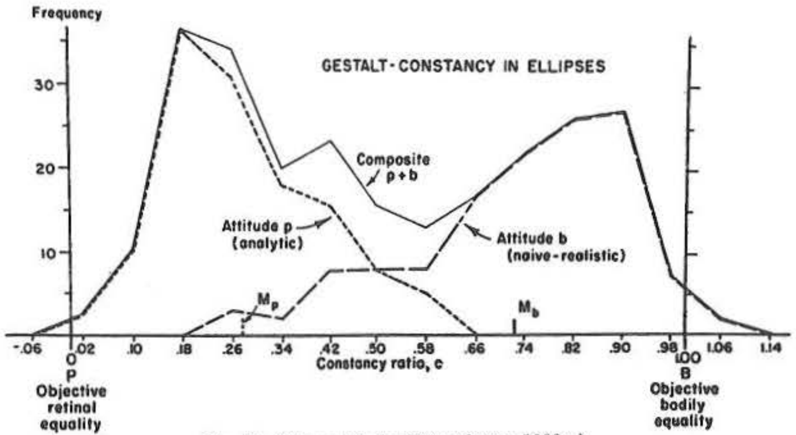


Fig. 33. (From data by Klimpfinger, 1933*a*.)
Distinction of Attitudes Within Perception in an Experiment on Shape Constancy.

with practice (see fig. 29), may be taken as an objective, statistical criterion for the existence of two distinctly different ways of perceptual field organization or stimulus evaluation.

One may even go one step further. The right-hand component curve is skewed to the left. It was established that this holds to a statistically significant degree (p better than .001). Such skewness may be used, again without recourse to introspection, in an attempt to spot at least some of the instances in which there was no understanding or no obedience to the instruction in question. With skewness extending under the mode of the *p*-curve, we must assume that the analytic attitude was mistaken for the desired natural attitude, *b*, in the cases represented by the tail.

So long as only attitudes *b* and *p* are involved, we are clearly moving within the framework of perception proper. We now turn to another experiment (for brief advance report see Brunswik, 1948) in which critical (or "betting") attitudes calling for the coöperation of the intellect proper were also involved; in addition, different avenues for the piping in of relevant information were used so that the effect of a compounding of perception and abstract fact-gathering could be studied in some detail.

As described in figure 34, an experimental laboratory class did the Müller-Lyer experiment first in the ordinary natural whole-perceiving attitude

toward length (*l*). As in all parts of the experiment, the series started from settings opposite to the expected illusion (negative values in the figure) so that any possible effect of "adaptation level" would work against rather than in favor of the illusion. Next, a poll of the results was taken and an evaluation given. Group A was then told about the average magnitude of the

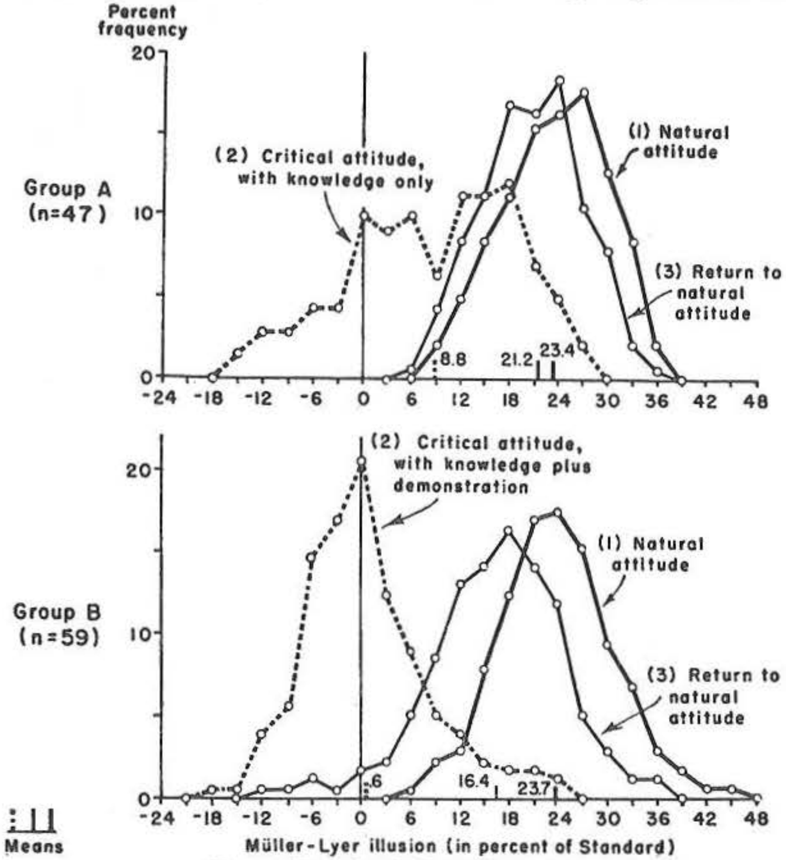


Fig. 34. (From data by Brunswik, 1948.)
Perceptual Attitudes and Interaction of Perception and Thinking in an Experiment on the Müller-Lyer Illusion.

illusion in abstract, numerical terms only by giving the mean per cent deviation of the group-PSE from the Standard (23.4, as indicated in the figure). Group B also was told its own group error (23.7, very close to that of group A); but in addition it was exposed to a perceptual demonstration of the correct setting and of the mean illusion setting, vividly supported by actual measurement of the correctnes of the former and the incorrectness of the latter. This measurement was performed in front of the class and quieted even the most tenacious doubters.

For both groups there followed a further experimental series in critical or "betting" attitude (2), asking for deliberate stimulus error in the sense of a superimposition of intellectual control over perception proper. As becomes clear from the broken curves in figure 34, group A retained a sizable illusion (mean = 8.8) in an obviously unharmonious, bimodal distribution which may be taken to reflect underlying conflict between perception and abstract knowledge as two distinct cognitive functions. Group B, on the other hand, with additional memories of the perceptual unmasking of the illusion to support its abstract knowledge of truth, was able to integrate the latter more fully with perception proper. On the average it almost fully compensated for the illusion (mean = .6). The critical ratio between groups A and B on (2) is 2.87. In addition, the distribution for group B shows remarkable unimodal clarity in spite of its skewness extending under the modal area of (1) in a manner similar to figure 33.

A return to the original, naïve-realistic attitude, required as an additional judgment on each presentation of the critical series, was also induced in the subjects. Results are presented in the thin solid curves (3). A comparison of groups A and B shows that such a return is—to a statistically significant degree (CR = 2.75)—less feasible for group B, for which the natural illusion had been upset not only by the relatively alien intellectual information but in addition by the more closely allied process of actual perceptual measurement.

3. THE AUTONOMY OF PERCEPTION: PERCEPTUAL AND CRITICAL ("BETTING") ATTITUDES AND THE STIMULUS ERROR

The relative functional "autonomy" of perception as held against the intellectual, measurement-calculational approach to reality (see Brunswik, 1934, pp. 118 f., 127) has been amply demonstrated in the past. It is evident not only in all cases in which intuition and "immediacy" are in conflict with "better" knowledge and wisdom (not always better in fact, as we have seen), but is also brought out experimentally. Examples are furnished, among others, by the more or less pronounced persistence of the perceptions of social objects or of the various geometric-optical illusions in the face of knowledge to the contrary, the relatively slow acquisition or extinction of new perceptual cues with artificial changes in the ecology (see § XVII) and by the attendant phenomenon of "learning without awareness." The functional discreteness of perception and thinking has also been brought out by our experiments reported in subsections 1 and 2 of the present chapter.

Yet the term "relative" must be stressed in speaking of the autonomy of the cognitive functions. They are to each other not like one planet is to another, but rather like the left and the right bank of a city, separated by a stream yet connected by bridges if only at certain points and for a limited volume of traffic. The experiment on the Müller-Lyer illusion reported in subsection 2 was designed to bring out some of this cross-cognitive traffic.

The distinction between cognitive functions of different levels as well as that between attitudes within the same function has traditionally been in a

bad way in psychology, however. In Part One we have noted the fact that even some otherwise outstanding investigators have been more than casual about inducing the proper attitude in constancy experiments or have failed to state clearly which attitude they intended to induce. The example given on page 23 concerning lack of clarity even as to the pole-variable used in evaluation is but a particularly drastic case out of several that could be quoted.

The notion that any departure from the most natural, or "naïve," perceptual attitude implies leaving perception altogether in favor of the intellectual approach and thus implies a "stimulus error" is widespread, even in otherwise sophisticated texts. Wundt's concept of stimulus error has been renamed "object error" by Titchener and defined as attending to and reporting the object or the objective fact, rather than the (proximal) stimulus or the simple sensation that was assumed to correspond to the stimulus. The concept is ill-conceived in view of its doctrinaire association with structuralism and the "constancy hypothesis" assuming a one-to-one correspondence between stimulus and sensation. In fact, under this point of view the natural *b*-responses could be—and have been—accused of being fraught with a distal stimulus error; while *p*-responses could be accused of a proximal stimulus error. In reality both are but different modes of structuring *within* perception proper while the stimulus error refers to an (illegitimate) intrusion of intellectual processes.

In order to help clarify these matters, we reproduce here the instructions given in our survey of size constancy (1944, p. 4 f.) for the two purely perceptual and the two part-intellectual ("betting" or critical) attitudes:

(1) *Naïve perceptual attitude, b:* Give your estimates on the basis of your first impression of the sizes of the objects in question. You should consider the sizes of the "things" as seen in the ordinary attitude of daily life (not projective sizes relative to your location). Do not let yourself be influenced by your abstract knowledge about, or memory of, the sizes of the objects in question, or of optics, etc.

(2) *Analytical perceptual attitude, p:* Try to perceptually analyze or to disintegrate the scene, in the way a painter would have to see it in order to be able to draw a perspectively correct picture—in other words, try to estimate visual angles, or the relative sizes of the projections of the objects as they could be measured on your retina or on a photograph with the camera set up where you are standing. Relate your judgment to an imaginary meter stick in a frontal plane at one meter distance from the eye. That is to say, the judgment should refer to the extension, apparently cut out on an imaginary frontal plane at one meter distance, by the light rays issuing from the object into the eye (but, of course, without the use of instruments such as a pencil held in front of the eye). As in attitude (1), however, you should rely exclusively on perceptual appearance after the field has become reorganized in the fashion described.

(3) *Realistic betting attitude, b′:* As in (1), concentrate upon "things." This time, however, critically superimpose upon perception all abstract knowledge available to you. Take the attitude you would have if you were to bet upon the sizes in question to the best of your knowledge.

(4) *Analytical betting attitude, p′:* As in (2), try to compare retinal (projective) sizes, again in relation to an imaginary meter stick at one meter distance. This time, however, take a betting attitude analogous to (3).

As can be seen, the explanation of the attitudes to be taken is not an easy or brief task and requires discrimination on the part of the subject. It may be aided by pictures demonstrating the various aspects of reality to be compared (such as fig. 70 in Gibson, 1950). It is advantageous to explain all the attitudes at the same time so as to show that only the betting attitudes involve an intrusion of intellectual processes (and do so on purpose). By actually inviting the stimulus error in these latter two attitudes we may help to prevent its occurrence in the first two.

The differences in the results of the purely perceptual and the corresponding betting attitudes bespeak the success of keeping thinking out of perception at least to a certain extent. Table 3 brings together all the relevant results cited in this book. It shows that even perceptual attitudes already focused on a certain variable may be further differentiated—such as in the

TABLE 3

PERCEPTUAL AND CRITICAL ("BETTING") ATTITUDES AND THE DEFINITION OF THE "STIMULUS"

[Note that for the Müller-Lyer experiment fringing area (F), being more global than length (L), has been paralleled here with the distal variable (B) rather than with (P). For the last of the experiments listed the F-pole has not been specifically ascertained.]

Functions or attitudes		*Size constancy* Holaday (See p. 22) c-ratios (log)	*Size constancy* Laboratory demonstr. (See p. 22) c-ratios (log)		*Müller-Lyer illusion* M. Müller (See p. 16 f. and fig. 2) mm	*Müller-Lyer illusion* Brunswik (See fig. 34) Per cent error
Purely intellectual (Measurement-computational definition of the "stimulus")	B	1.00	1.00	$F_{corr.}$	47	
Critical ("betting"): Perceptual superseded by intellectual	b'	.94				
Purely perceptual	b	.86	about .9	f_i	51	
				f_w	54	
				d	65	
				l_w	100	A: 23.4 (21.2[a]) B: 23.7 (16.4[a])
Purely perceptual	p	.39	about .6	l_i	112	
Critical ("betting"): Perceptual superseded by intellectual	p'	.25	about .2	l'	126	A: 8.8[b] B: .6[c]
Purely intellectural (Measurement-computational definition of the "stimulus")	P	0	0	L	150	0

[a] After betting attitude as below.
[b] With knowledge.
[c] With knowledge and perceptual demonstration.

Müller-Lyer experiment reported in Part One—and that the superimposed intellectual function may either be left relatively speculative or reinforced by specific information; this latter may in turn be further reinforced by the memory of the perceptual impact of a measurement demonstration (see the discussion of fig. 34 in § XIV/2). The purely intellectual, measurement-computational approach at the same time constitutes the definition of the "stimulus" (or of the respective poles of intention, designated by capital letters throughout our presentation). So long as estimates deviate from accurate stimulus values, any existing "stimulus error" is likely to have been mitigated by a genuinely perceptual component, at least so far as the experiments reported in table 3 are concerned.

The fact, mentioned on page 22, that betting attitudes tend to cut error approximately in half (or even in one-third) relative to the intended pole, and thus demonstrate the presence of a quasi-autonomous contribution from outside of perception, is well brought out throughout the table.

We must add that this finding seems not to be ecologically generalizable beyond the visual area, even if it could be proved to hold generally within that area. Mohrmann had introduced the betting attitude in some special series of his study on loudness constancy (1939, p. 184; for main report see above, § XI/2); his results were inconclusive on this point, pointing even to a slightly disruptive effect of intellectual entanglement that did not reach statistical significance, however. The possibility of constructive integration of perception and the intellect thus remains in suspense so far as the less highly organized or the lower sense departments are concerned.

XV. The Study of Physiognomic Perception by Systematic-Representative Hybrid Designs

As do many things alive, science grows at its seams. Rather than by the pursuit of the extreme or pure paradigm, progress is often carried by the straddling of lines. In some cases hybridization is merely a matter of intercombination in the sense of reapplication, such as in representative design which grafts the controlled uncontrol of statistical sampling upon the hitherto rigidly controlled content of experimental psychology.

Yet, by fully espousing the statistical principles, representative design itself constitutes an extreme case as to form. In this capacity it becomes capable of hybridization in a second sense of the word, one that is more nearly a case of compromise or of intercombination between basic principles.

The examples which we will discuss in the present chapter are in the no man's land between pure systematic and pure representative design, some of them closer to the former and others closer to the latter. They involve both combination and compromise between the two, and they do so in a variety of ways and to different degrees. All of them are concerned with "physiognomic" perception, if this term is allowed to stand for impressions of person-like, animated qualities elicited not only by faces or face-like drawings but also by representations of the full human figure. The question is also posed as to how far down we may go on the ladder of resemblance to actual human appearance without losing the physiognomic quality of impression. The farther we can go, the more we may believe in the distal aim of perception. In the case of physiognomic perception this aim extends in "depth of intention" (see Brunswik, 1934) beyond the overt distal region as covered by the ordinary perceptual constancies and into the covert distal region (see above, fig. 7).

We will begin with an elaboration of the schematized face experiment by Brunswick and Reiter (1937) briefly referred to in Part One (p. 42) and proceed through various intermediate steps to an experiment by Wallace (1941, 1947) in which artificially distorted photographs of actual persons were used. The schematized-face experiment is a relatively systematic experiment and therefore is largely confined to what we have called the impression problem (pp. 29, 33, 50). This problem is concerned with the utilization of stimulus patterns by the responding system regardless of the correctness of such utilization in the particular or the general case. In other words, it is concerned with the functional correspondence of the response relative to the proximal (or overt-distal) but not relative to the (covert-)distal stimulus. The distorted-photographs experiment, on the other hand, is relatively close to representative design and therefore more capable of interpretation in terms of distal achievement along with impression value. A number of other sequel studies of the schematized-face experiment will also be brought into the picture. Some of them will serve to illustrate the trend, characteristic primarily of the differential-psychological and testing tradition of this country, to build up the size or variety of the responder-populational sample (n) at the expense of the size or the variety of the ecological sample (whether this be a representative sample, N, or a mere multitude, M; for the use of these letters see below).

1. TRUNCATED FACTORIAL DESIGN IN STUDYING THE IMPRESSION VALUE OF SCHEMATIZED FACES

At least tacitly, any physiognomic experiment envisages some kind of appraisal of distal achievement as one of its ultimate ends and thus is bound to the principles of representative design. Yet, the designing of the Brunswik and Reiter experiment (1937; see above, p. 42), begun in the early 1930's, was in the main still dominated by the ideology of the systematic experiment. In certain respects, however, the design is tacitly guided by principles of representativeness. The original aim was to investigate the impression value of a number of prominent features of the facial geometry including their combination in patterns or "Gestalten." In the selection of these variables some consideration was given to their possible ecological validity, but this was done only informally. Actual photographs of real persons seemed unacceptable because of the large number of uncontrolled factors, a feature which under the auspices of representative design would be considered an asset rather than a liability. Only a schema seemed capable of offering the desired control. It would permit the systematic variation of a limited number of factors, other factors either to be omitted or held constant. A first impasse developed at this point. It turned out that when constant features were added, such as hair style in a certain manner, all faces tended

to look alike; on the other hand, the omission of too many constants, such as of the pupils in the eyes, tended to create a mask-like impression with much of the physiognomic quality lost. Both of these shortcomings are penalties for departing from genuine representativeness in the design.

The standard or Norm face is shown in figure 35. It was arrived at after considerable preliminary trial and error and represents the best compromise in our attempt to avoid the fixation of the impression quality, on the one

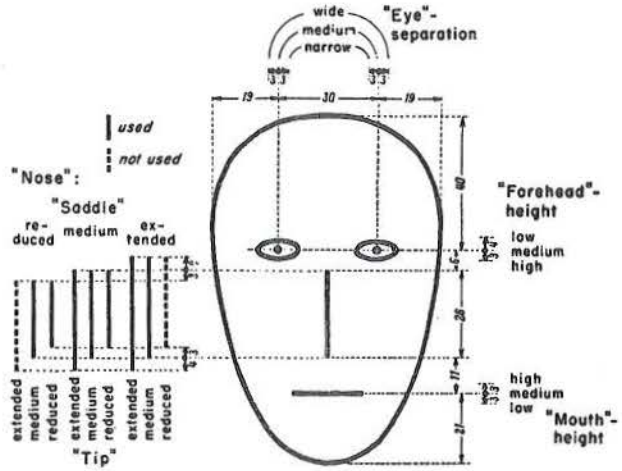


Fig. 35. (After Brunswik and Reiter, 1937.)
Factorial Variation of a Schematized Face.

hand, and its impoverishment, on the other. In addition, it was selected among many of its variants as the most "normal" looking—another concession to representativeness. Still another decision made with an eye on representativeness was the choice of three levels of strength for each of the experimental variables so as to enable the design to bring out a suspected curvilinearity. This choice is shared with many factorial designs, to be sure. In our case there was the added note that a "normal"-looking standard was used in defining the center position. We laid stress on this feature of the design because it made possible an investigation of the old problem of the more favorable position, in physiognomic matters, of the mean as contrasted with the extremes.

The number of variables eventually decided upon was dictated mainly by considerations of practicability, that is, by the limitations in the number of intercombinations that would still allow effective comparisons by the subjects (fig. 36). We tried to go as high as possible in this respect. Indeed, the experiment was influenced by the multipolar designs in thing constancy which we have discussed in §§ XI and XII, and was seen as an investigation in multidimensional psychophysics. The use of all (or rather, as we will see in a moment, of nearly all) intercombinations of several component features seemed a rather obvious procedure in view of the fact that in a face one may easily find some factors that are independently variable while in our multipolar weight and volume experiments the factors considered had been rather interdependent.

This brings us to a peculiarity of our design which is of considerable theoretical importance in precisely locating this experiment on the map of possible designs. It is given by the fact that two of the nine "noses" that would result from the use of all intercombinations of three "saddle" with three "tip" positions (fig. 35) turned out so unrealistic and grotesque that they threatened to spoil the seriousness of an already precarious attitude on the part of the subjects. It is at this point that our ignorance of Fisher's work on factorial design and its mathematical evaluation (1925, 1935) paid off. Unfamiliar with the advantages to be gained from an application of the analysis of variance to the results of a factorial design provided that it is complete, we omitted the two obnoxious noses (broken lines in fig. 35) for the sake of improved representativeness.

Our design is therefore not a factorial design in the strict sense; it may perhaps be called a truncated factorial design. It is best visualized from figure 38. Instead of filling all nine cells, as we have done in the cases of the intercombinations of "forehead" with "mouth" with "eye," the entries are confined to a slanting area that leaves out two opposite corners. The relationship of the present truncated case to representative design proper becomes striking by a comparison of figure 38 with the design for our fully representative survey of size constancy presented in figure 8. Both show the natural obliqueness of imperfect relationships rather than the complete independence of factorial design or the perfect dependence of classical tied-variable design (fig. 4, A), even though in figure 38 a strange orderliness is superimposed upon the naturalness of the bivariate distribution which characterizes figure 8.

On superficial scrutiny factorial design may appear ideal since all plots are filled while in representative design normally only part of the plots are filled; representative designs thus would seem readily extractable from factorial designs if anyone should wish to extract them, but not vice versa. The first catch is, as we have seen, that some of the intercombinations of variates may be incompatible in nature or otherwise grossly unrealistic. Of more basic importance, at least in some cases, is the fact that for the proper functioning of cue-object and means-end relationships existing ecological validities must be left undisturbed lest a distorted picture of psychological functioning be created.

Lumping the remaining seven noses together, our design yielded $3 \times 3 \times 3 \times 7 = 189$ facial schemas. The faces, together with their arrangement and designations, are presented in figure 36. The Norm appears in position H5.

All faces had been printed by means of an adjustable press so that the identity of the constituent parts was maintained as closely as possible. The vertical diameter of the facial oval was 104 mm. The entire table was mounted on a well-lighted wall faced by the subject.

According to a broadly conceived principle of representativeness the full range of physiognomic attitudes or response dispositions elicited by the schemas would have to be surveyed before deciding on the qualities of response to be asked of the subjects. A step in this direction has been taken when in exploratory demonstrations a limited number of schemas was shown to casual observers and their free descriptions were invited. These descriptions served as a guide in the final selection of the qualities to be judged. Some of the free descriptions were:

H5 (Norm): Normal
P1: Narrowminded.
G9: Joyful, openhearted, young.
T7: Old, embittered.
P3: Soft, sad, an intellectual.
D9: Sarcastic, calculating, a merchant.

(The reader will find enlargements of these faces in Brunswik, 1934, as also reproduced in Allport, 1937.) The seven qualities chosen for the main experiment are shown in figure 37.

Ten educated adults served as subjects. Only a single row or column was exposed at any one time. The subject was asked to rank each row and each column according to one of the impression characteristics or physiognomic qualities, the order of rows and columns being randomized in a different manner for each subject. After multiplying all the row ranks by 3 and all the column ranks by 7 to give them equal weight, the horizontal and vertical rank-scores were added for each face and the sums converted into rank-indices ranging from 0 for least favorable to 100 for most favorable. The ranking procedure was repeated for each of the seven qualities.

As was the design, our original evaluation was makeshift. By averaging the rank-indices obtained from the ten subjects for each face and each quality separately, mean rank-indices were obtained. For apparent intelligence, these mean indices appear in *italics* at the lower right of each face in figure 36. The mean rank-indices were then averaged in blocks, much as is done in preparing for an analysis of variance. Thus, within the first three groups of three columns each, any single point represents one-third of all facial schemas, that is, 63, whereas any point in the last group of seven columns represents one-seventh, that is, 27, schemas.

In figure 37 the qualities are listed not in any predetermined order but in accordance with the greatest similarity as to the over-all graphic trends in the respective curves so as to bring out their underlying affinity in the response system of the raters. Concerning the affinities between the various impression qualities in terms of the external conditions it is fairly clear from figure 37 that apparent mood and age tend to go together, while character, likability and beauty form a second group, and intelligence and energy a third. Perhaps this can be set in parallel to the halo-effects mentioned in connection with Experiment D in which photographs of real persons had been used (see p. 28 ff.).

Turning now to the dependence of the various physiognomic qualities upon the geometric features varied in the schema we note that variation of

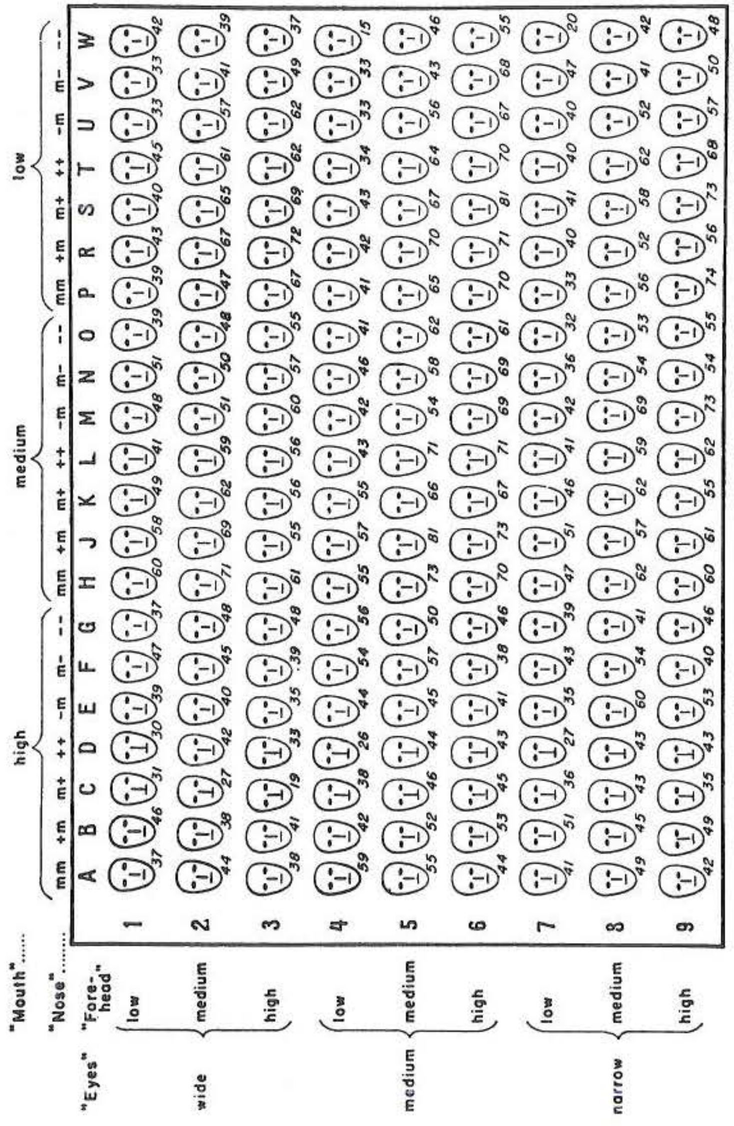


Fig. 36. (After Brunswik and Reiter, 1937.)
Schematized Face Experiment, with Indices for Apparent Intelligence.

"mouth" elicits the most extreme responses, high mouth (chin) appearing as gay and young, low chin as sad, old, and bad. A note of ambivalence is injected by the fact that high chin, while favorable in some respects, swings down to unfavorableness as we proceed to apparent intelligence and energy. Wide eyes and short nose exert influences somewhat similar to those of high mouth even though to a lesser extent. The longest noses are unfavorable

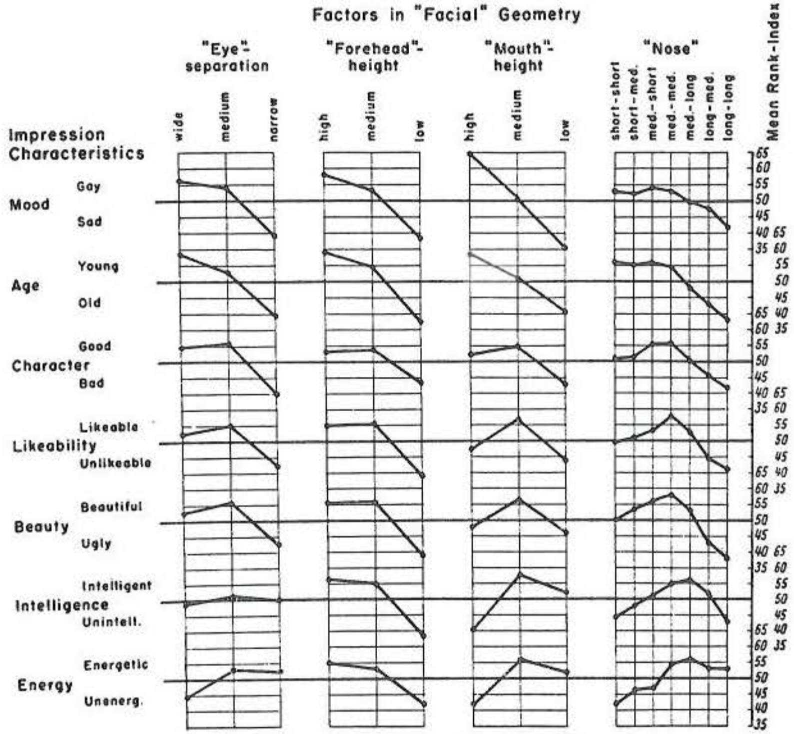


Fig. 37. (After Brunswik and Reiter, 1937.)
Mean Rank Indices for Four Factors Varied, on Each of Seven Impression Characteristics.

throughout. "Forehead" or brow, while likewise not differentiating between qualities, is uniformly favorable when of the high and medium variant. "Eye-separation" is almost without effect upon apparent intelligence. And for all facial features the medium variant tends to be on the favorable side, reminding us of the "golden mean" so celebrated in antiquity.

S.D.s for each of the 189 mean indices were estimated for apparent intelligence and were found to average 5.6. On the basis of this informal appraisal and with certain additional assumptions, differences in figure 37 may reach statistical significance when surpassing a few scale units.

The analysis of "interaction" among facial features was left to still more informal means of evaluation than the study of the main effects reproduced in figure 37. The attempt was confined to two relatively haphazard modes of approach. One consisted in the grouping together of some specified configuration patterns. This showed, for example, that "narrow" faces tended to appear especially sad, old and bad, "broad" faces especially happy and young, while faces in which the features were spatially far apart tended to elecit negative responses all along the line.

The second approach to interaction consisted in a study of those individual schemas which incorporated in themselves the ideal additive combination favoring a certain impression. The presence of interaction was revealed by the fact that the expected effect did sometimes not even nearly come off. This was found true especially for intelligence and energy. Thus K6, a combination of all the extremely intelligent-appearing features (as is revealed by comparing fig. 36 with fig. 37) was found to occupy but 25th place on the list of rank-indices; even more drastically G1, the ideal additive combination for unenergetic, is 90th from the bottom of the list for energy, that is, occupies a place just about in the middle of the list.

2. AN IMPASSE IN THE APPLICATION OF THE ANALYSIS OF VARIANCE

It goes without saying that in the appraisal of the significance of the main effects, and of the presence of interaction as such, the analysis of variance furnishes a much more powerful tool than the semi-technical methods which we have used. But aside from the fact that informal methods may offer a more ramified picture of the Gestalt-type interactions with which we are here concerned, analysis of variance is confined to complete factorial designs and breaks down even in those modest digressions toward representative design which is given in our case of a truncated design for the two "nose" variables.

We may nonetheless carry out a routine analysis of variance[21] for the factorially orthodox part of our experiment, that is, for the variables "forehead"-height, "mouth"-height, and "eye"-separation. Since only the mean rank-indices as shown in figure 36 were available, higher-order interactions had to be taken as the error term. Second-order interactions, all of which had been found below significance in a preliminary analysis, were included in the residual. As can be seen from table 4, all the main effects and part of the first order interactions are statistically highly significant. The order of importance for the first three variables is (1) "forehead," (2) "mouth" and (3) "eyes." No formal test for the statistical significance of this order has been attempted, but an inspection of figure 37 will show that it agrees well with the relative steepness of slants in the graphic picture for these three variables so far as the rubric "intelligence" is concerned.

Turning now to the unorthodox, truncated part of the design which involves the two "nose" variables and which is, as will be remembered, the result of a compromise toward representativeness, we find ourselves in an

[21] This analysis, as well as that for the truncated part of the design reported below, was undertaken, as part of a joint advanced course in experimental psychology and design, by Professor Rheem F. Jarrett with the help of Dr. Robert Rollin. The analysis uses the mathematical model or version in which constant values for the variables are assumed. No technique seems as yet to have been developed for a test of the main effects against the significant first-order interactions with more than three variables in the alternative mathematical model known as variance-components analysis; for beginnings see Crump (1951) and Cochran (1951).

impasse so far as the applicability of the analysis of variance is concerned. Analysis of variance demands that all plots be filled while in this part of our design, as we have seen, two opposite plots are unfilled (fig. 38). Statisticians provide two ways out which may be applicable to situations like ours, even

TABLE 4

ANALYSIS OF VARIANCE OF INTELLIGENCE JUDGMENTS FOR THE ORTHODOX FACTORIAL PART OF THE SCHEMATIZED FACE EXPERIMENT

Source of variability	Degrees of freedom	Mean square	Variance ratio F
Main effects			
"Forehead"-height	2	3,750.04	116.17[b]
"Mouth"-height	2	3,107.44	96.27[b]
"Eye"-separation	2	604.00	18.71[b]
"Nose"	6	353.75	10.96[b]
First order interactions			
F × M	4	874.11	27.08[b]
F × E	4	126.02	3.90[b]
F × N	12	52.41	1.62
M × E	4	89.02	2.76[a]
M × N	12	279.00	8.64[b]
E × N	12	48.60	1.51
Residual (second and third order interactions taken as error)	128	32.28	.....

SOURCE: Computed from data by Brunswik and Reiter, 1937.
[a] Statistically significant between .05 and .01 levels.
[b] Statistically significant at .01 level or better.

TABLE 5

ANALYSIS OF VARIANCE OF INTELLIGENCE JUDGMENTS FOR THE TRUNCATED PART OF THE SCHEMATIZED FACE EXPERIMENT

Features of "Nose" as sources of variability	Degrees of freedom	Mean square	Variance ratio F
"Saddle"	2	361.68	11.20[a]
"Tip"	2	807.86	25.03[a]
Residual	128	32.28	.....

SOURCE: Computed from original data by Brunswik and Reiter, 1937.
[a] Statistically significant at .01 level or better.

though not without some intellectual trickery. One is the "missing plots" technique which would require the pretense that part of the experiment went wrong or that some of the data were lost but really should be there. In our case a second alternative was used, equally makeshift so far as application to our case is concerned, to wit, the so-called "unequal numbers in classes" technique (Kendall, 1951, Vol. II, p. 220 ff.). Zero frequency and an undecided score were assumed in the two unfilled plots in figure 38. As can be seen from table 5, both "saddle" and "tip" have highly significant main effects. The present technique does not provide a test of interaction for the semi-representative type of design analyzed here.

The impasse just described illustrates the need for a broadening of the analysis of variance so as to legitimize truncated designs, or else for the utilization or development of other methods of evaluation which would take care of fully representative designs also. We have already refuted the possible argument that the artificial filling up of empty plots or areas in a natural design is feasible in the general case or is harmless or even beneficial.

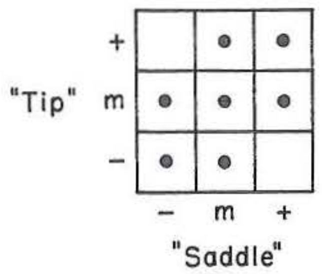


Fig. 38.
Truncated Factorial Treatment of "Nose" in Brunswik and Reiter's Schematized Face Experiment.

3. THE GENERALIZABILITY OF THE RESULTS AND ALTERNATIVE MODES OF EVALUATION

It remains to discuss the important problem as to what exactly is meant by the "statistical significance" of our results. It will be remembered that in Part One (p. 38 f.) we have distinguished between responder-populational and ecological types of generalization and of significance. The latter requires the presence of an ecological universe from which the stimuli have been sampled. But in the case of our schematized-face experiment there is no such reference class; our 189 drawings therefore do not merit to be labeled an N, that is, a sample in the representative sense; they constitute no more than an M, that is, a mere multitude of cases with a highly questionable status as to what they may be able to stand for.[22] At best, an M is an artificial universe in itself, but this rechristening will not add to its generalization potential. Ecological generalization, in the technical sense of the term in which we would like to see it applied, can be made only on the basis of an N, never on the basis of an M. Whatever quantified significance our results may be able to claim is purely responder-populational in character, referring in our case in large measure to incidental fluctuations along intra-individual

[22] Note that side by side with N the capital letter M is used here instead of the lower-case letter m introduced on page 36 f. This slight revision in nomenclature has been decided upon so as to achieve uniformity in the use of capital letters for all ecological referents. Correspondingly, n and m are henceforth to refer, respectively, to true samples vs. mere multitudes of respondents (or respondent-conditions).

dimensions or a-variables (pp. 6, 33) which must be held responsible for much of our error term.

We must therefore face the fact that such relatively minor modifications as a change in the shape of the "mouth" or the addition of hair may, in principle, affect the result rather markedly. In fact, all the things that were omitted or held constant in the experiment add to this uncertainty of ecological generalization. This means that any expansion of our results to live faces could be done on the basis of non-technical plausibility arguments only.

It is not even possible to say with precision how far the various lines and patterns in our schema mean what they purport to mean. For this reason we have made a point of always setting the respective terms in quotation marks. For example, whether "forehead" is taken to include the area to the hairline only or whether it means the full oval including the hair cannot be determined offhand. In our original publication we have speculated about such questions at some length on the basis of our results; in our particular example we have come forth with the suggestion that since low "forehead" creates the impression of age rather than of youth and older persons are more frequently bald, our "forehead" probably includes the hair. That "nose" could not stand for the anatomical nose has already been mentioned on page 42 in connection with the follow-up study by Samuels. Since furthermore a reduced "tip" makes for the appearance of youth we may perhaps conjecture that "tip" really stands for tip rather than for the root of the nose (its intersection with the upper lip). This type of reasoning makes informal use of facts of ecological validity (e.g., hair as a cue to age) and thus touches upon the distal achievement problem; but it must be stressed that all this is done outside the domain of statistical evaluation proper.

Compare this with the case in which each of the 189 schemas would be replaced by a photograph of an actual person so selected as to match the features of the respective schema. This would be nothing but an expansion, from 10 to 189, of the procedure already employed by Samuels (1939) in the second of her two experiments (for further detail see below, subsection 4). If the selection of the faces is otherwise fairly random—as it almost would have to be for practical reasons—another source of error variance is introduced which is of an ecological kind rather than being due to accidental inter- or intra-individual fluctuations. Although the new collection of stimuli still contains elements of systematic variation and thus 189 remains an M rather than becoming a true N, certain ecological generalizations ("applications") or comparisons with previously non-varied variables now become somewhat more defensible. For example, if Samuels' contention that our main effects are reduced to secondary importance when photographs are used instead of schemas, should be borne out for the larger M, we would have to accept her criticism that we had not chosen the most influential aspects of facial geometry for our systematic variations; at least this would be reasonable to conclude if in addition it could be shown that the levels of strength for our systematic variables were fairly evenly distributed over the range of natural variability of these variables.

With this latter stipulation in mind, it becomes questionable whether the hybridization of systematic and representative practices which would take place in our fictitious experiment would have much of a point left. There seems no cogent reason why not to proceed to an experiment of the type of Experiment D as described in Part One. Such an experiment employs representative variation equitably for all the variables of the facial ecology and not only for the "left-over" variables. A real and accurate test of the relative utilization strength of the different ecological factors cannot be obtained in any other way.

The analysis of variance is, at least in its present form, not a very suitable tool in the evaluation of representative data of this kind. A better way would be to proceed in the general manner of figure 6 and the adjoining text throughout § VI. The analysis was there not pressed beyond simple correlation; since both the impression values (utilizations) and the ecological validities were low even for the best of the facial cues we could find, no attempt was made to establish the rank order of their utilization strengths by the use of technical significance criteria. But it will be noted that both Nose-height and Forehead-height had stood out as comparatively valid in figure 6, and since these two cues also appear in our schematized-face experiment, Samuels' criticism of our selection of variables loses a good deal of its cogency.

Further mathematical steps in the evaluation may perhaps be patterned after a model first introduced into the study of functional problems in perception by Else Frenkel-Brunswik in her analysis of clinical intuition (1942; 1951, p. 362), that is, multiple correlation. For facial cues an example of such an analysis of impressional values was undertaken from the materials underlying our figure 6, using Nose-height (N) and Forehead-height (F) in relation to the response quality, likability (l). The intercorrelation between N and F was ascertained as .27, and their respective impression validities relative to likability as .30 and .28. The multiple correlation between Nose-Forehead and apparent likability is then computed as $R_{l \cdot NF} = .41$. Prediction of response even from low-utilization cues could be built up to a considerable extent by such a procedure. And the same procedure could be used symmetrically for ecological validity so that eventually an understanding of the mechanism of the superordinate functional achievement arcs (correctness, fig. 5) could be obtained.

What, then, is left to do for factorial or other systematic experiments along this line? Unless one is willing to go into plausibility generalizations—always precarious in nature—or is satisfied with results confined to a self-created ivory-tower ecology, the answer is: not much, except an exercise in neatness. That there may be rewards in the field of differential-psychological problems, and that the struggle to cross the no man's land between systematic and representative design may be a challenging process of tentative enrichment and approximation in spite of the lack of rigorous generalizability is the theme of the remainder of this chapter.

4. DIFFERENTIAL-PSYCHOLOGICAL EXPANSIONS OF THE STUDY WITH SCHEMATIZED FACES

The further studies using our schematized faces or their derivatives may be divided in two groups. In one the emphasis is on differential-psychological or comparative generalization or specification; the number of subjects, n, is increased, and sometimes normals and the mentally diseased or adults and children are compared. Usually this goes with a drastic reduction in the number of faces, M. In the other group of studies the emphasis is on stimulus expansion, that is, on increased ecological representativeness in the sense of a more varied M.

In the former group belongs the first part of the study by Samuels (1939, pp. 20–22), mentioned earlier in this chapter. Working under the direction of Gordon Allport, Samuels used ten faces that had been singled out by Brunswik and Reiter (1937, p. 128) as extremes on the impression scale, both desirable and undesirable. These faces are reproduced in figure 39; they are identifiable in their geometrical composition by their row- and column-designations as given in figure 36. Their positions from the top or bottom of the lists of mean rank-indices are given whenever especially extreme. Samuels' purpose was to check our results on a larger sample of 247 Harvard and Radcliffe students, thus reversing the n/M ratio from our original 10/189 to 247/10 (see also above, p. 42). The faces were redrawn and presented in paired comparison.

These changes made possible the establishment of the inter-individual generality of a crucial part of our results, a purpose for which we have had too few subjects. Samuels selected one pair of opposites for each of the seven impression qualities. These are indicated in the left part of table 6 which also shows her results. The per cent agreement with the trend in our data is high. It ranges from 83 per cent "correct" (in terms of our Vienna findings) for the pair selected for likability to 96 per cent for the pairs for mood and beauty. The average is 88 per cent (as against a chance expectancy of 50 per cent). There are no sex differences; the average for the 184 men in the sample is 88 per cent and for the 63 women it is 89 per cent.

The same trend toward greater responder-populational variety while reducing ecological variety is evident in another of the American studies utilizing our schematized faces, one by Halstead (1951*a*, 1951*b*). While Samuels had used normal college students, Halstead attempted to develop the schemas into a test for brain damage or psychosis. He used only nine of the ten faces shown in figure 39. Omitted was S6, and the remainder were presented simultaneously in a three-by-three square arrangement. One face was to be selected for each of 14 characteristics, the desirable and the undesirable alternative on the seven impression qualities we had selected. Face J9 yielded no preference and remained unclassified. The remaining eight faces were classified in a simplified manner merely as to desirability. This is shown in the right-hand part of table 6 which also presents the results. For the desirable faces the chance expectancy of being "correct" (in terms of our results) is 33.3 per cent, and for the undesirable, 50 per cent.

Halstead reports (1951*b*, p. 124 f.) that his 158 brain-injured cases, 103 psychiatric patients, and 102 normal controls showed "striking similarities rather than . . . differences in the grouping of behavior." At the same time

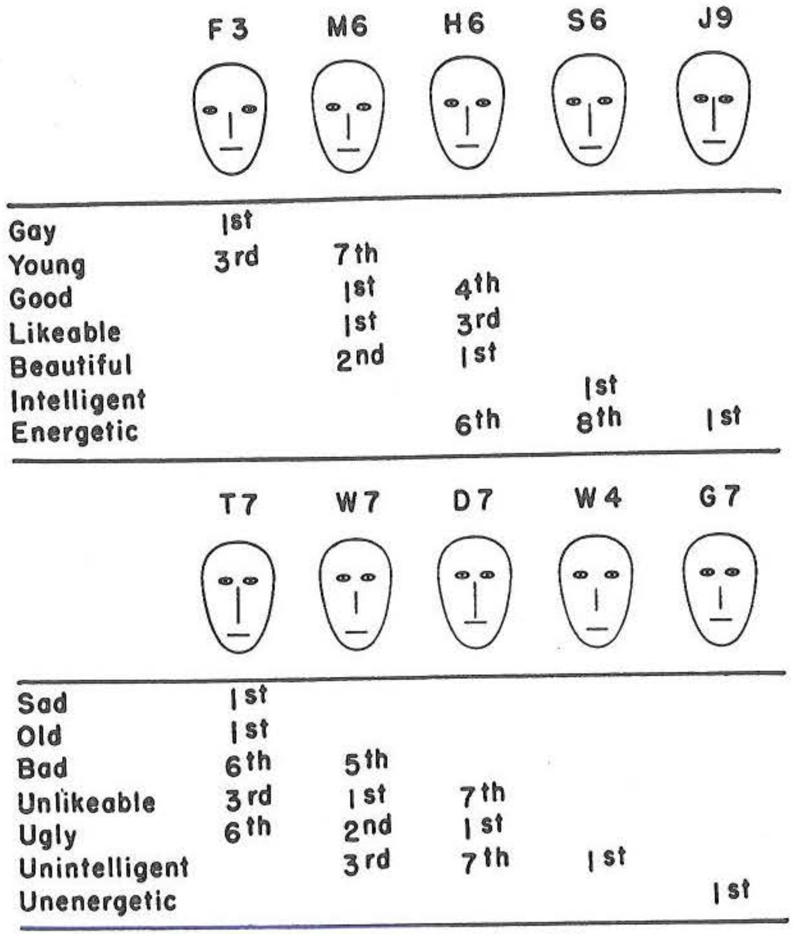


Fig. 39. (After Brunswik and Reiter, 1937.)
Schematized Faces Yielding Extremely Desirable and Undesirable Impressions.

the results are reported to agree with our original results in an average of 80 per cent (against 33 per cent chance) for the three clearly "desirable" out of the nine faces selected, and of 86 per cent (against 56 per cent chance) for the five clearly "undesirable" faces. (The agreement might have been still higher if the full differentiation of the original classification of the faces in terms of the seven impression categories had been maintained.)

Halstead adds that "we are not clear when this process begins, but we have found one instance in which a child of four years and ten months required but slight translation of the trait names to show this conformity or stereotypy in thirteen of the fourteen judgments." In terms of the above agreement scores, this would correspond to a 93 per cent agreement.

The physiognomic impression values of our schemas thus seem to be a responder-populationally highly generalizable characteristic, one that maintains itself even in the

TABLE 6

RESPONDER-POPULATIONAL EXPANSIONS OF THE SCHEMATIZED FACE EXPERIMENT

Impression qualities	Samuels (1939)			Halstead (1951*a*, 1951*b*)							
	Faces used		Per cent "correct" judgments (247 *Ss*)	Faces used		Per cent "correct"					
	Desirable	Undesirable		Desirable	Undesirable	Brain disease (158 *Ss*) D	Brain disease (158 *Ss*) U	Mental illness (103 *Ss*) D	Mental illness (103 *Ss*) U	Controls (102 *Ss*) D	Controls (102 *Ss*) U
Mood.	F3	T7	96			86	92	90	87	86	95
Age.	H6	D7	84			92	91	80	93	92	92
Character.	M6	T7	84			77	81	80	76	83	76
Likability.	S6	W4	83	F3 M6 H6	T7 W7 D7 W4 G7	77	82	73	84	68	89
Beauty.	M6	W7	96			86	85	84	84	86	86
Intelligence.	F3	T7	89			87	85	91	89	90	89
Energy.	J9	G7	87			64	86	56	85	60	86
Mean.			88			81	86	79	85	81	88

face of severe brain damage and psychosis and that can be traced relatively far down on the developmental scale.

Developmental and comparative data were also collected by this writer some years ago with children and adolescents from predominantly white, predominantly Negro, Chinese, and Indian-reservation schools on the United States West Coast. The variety of our schematized faces was cut down from the full four-dimensional "checkerboard" arrangement of the original factorial design to a four-dimensional center "cross" organized about the Norm, H5. Thus there remained only nine different faces, the Norm and one on each side of the Norm along each of the four major axes. As a step toward stimulus expansion of potential ecological relevance a new dimension was introduced, "width of skull at the temples." The choice of this variable was prompted by Cleeton and Knight's (1924) finding that of many facial variables considered, width of skull seemed the most promising as to physiognomic-ecological validity (r about .30, even if only with their relatively small N of 30). On the basis of a preliminary survey of the results (which have not been adequately analyzed so far), there seems little reason to assume any momentous differences between the ages or racial groups which we have studied.

5. ECOLOGICAL EXPANSIONS OF THE STUDY WITH SCHEMATIZED FACES

We now turn to the other group of sequel studies of the schematized-face experiment, those stressing stimulus expansion. In this group belongs the second part of the study by Samuels (1924, p. 22 f. and fig. 2), the part to which we have given some attention in subsection 2. It involves the use of one near-precise match for each of our ten impression extremes, the matches being selected from fairly standardized actual photographs in college yearbooks. Sixty student observers have yielded a "correctness" value still somewhat above chance although sharply reduced when compared with the value for the corresponding schemas (from 88 per cent, see above, to 63 per cent).

Samuels' technique represents a real effort toward ecological rapport, but the author is wise in interpreting her results merely as reflecting on the soundness of our original choice of variables, that is, as an indication of "the relative lack of influence of the Brunswik and Reiter cues in the real faces." We have already commented on this criticism above, and have done so with qualified approval. Any more assertive evaluation, however, might involve a fallacy, referred to in greater detail in § XVIII, which Hammond (1948, 1954) has exposed in some social-psychological and clinical studies. In these studies no more than one or very few live representants per "person-condition" (such as sex) were used when the persons were part of the situation (such as examiners in clinical testing), and these ecological person-conditions were treated in the manner of systematic design as if they were isolable "physical" conditions. Severe interpretational fallacies involving ecological overgeneralization were the result. Similarly, as we have said, the technical appraisal of the ecological generalizability of the relative weights of the variables of our schema has not even begun unless a representative sample is used for each of the ten schemas or, still better, a genuinely representative N is distributed over part or the whole of our M = 189 drawings.

An intermediate step between our crude schema and a photograph is given by the type of dimensional enrichment offered in three recent studies from the University of Vienna as summarized and illustrated in a text by Rohracher (1952, pp. 145–148, figs. 31, 32). In each case the Brunswik and Reiter schematic faces were taken as the starting point of new variation.

The impression value of the eye region was investigated by Kremenak (1950). She varied the size and shape (thickness) of the eyebrows, their distance from the eyes, and the shape of the eyes—to represent the slit between the eyelids upon which Lersch (1932) has laid such great emphasis. There were four shapes of eyes, ranging from a very narrow ellipse to a near-circle; four shapes of eyebrows, horizontal, arched, slanting upward toward the side of the face, and slanting downward toward the side of the face; and three distances of the eyebrows from the eyes. The position of the eyes and the size, shape and positions of the nose and mouth were constant within our familiar standard oval-shaped outline. A total of 144 schematized faces were judged on nine pairs of psychological attributes by 50 judges who were evenly divided between the two sexes. The attributes used were: good, sincere, careless, extraverted, gay, intelligent, energetic, handsome, likable, together with their opposites.

Kremenak found that with her particular stimulus material the eyebrows had greater expression value than the shape of the eyes. The shape of the eyebrows had the most influence; the distance of the brows from the eyes was next. A face with eyebrows placed close to the eyes, the brows slanting upward toward the side of the face, was most often judged as bad and introverted. A face with arched brows and a wide distance between the brows and eyes was judged as unintelligent and extraverted. With certain configurations the results were close to unanimous.

The mouth area was studied by Winkler (1951). She varied the shape, the height, and the angle of incline of the mouth in 238 configurations with 50 subjects and the same dimensions of impression as Kremenak. Among her variables, the shape of the mouth was found to have the greatest influence on the expression value. A wide mouth-slit with a downward angle at the corners gave the impression of badness (maliciousness); a narrow mouth-slit denoted introversion; a wide mouth-slit with an upward angle at the corners led to judgments of carelessness.

The impression value of hairline and of grooming of hair, beard and mustache was made the subject of a study by Seiller-Tarbuk (1951). A total of 224 configurations was used. Within this material the way the hair was placed on the head had the greatest influence. Hair style (parted at side, parted in middle, without part, bald) and kind of beard or mustache had decisive meaning only for the attribute pairs, gay-sad and energetic-unenergetic. Hair low on the forehead gave the impression of unintelligent, bad, and unlikable, while a higher hairline (also bald heads) gave the impression of intelligent and good (compare this with the discussion in subsection 3); beardless faces were judged as more intelligent, good-natured and energetic than faces with beards.

The ecological generalizability of these results, both as to direction of influence and as to relative strength of factors, is subject to the same limitations which we have discussed above for those of the major features of facial architecture which were varied in our own study. The crux in all these cases is that human appearance, and especially the face, constitutes as tight a package of innumerable contributing variables as might be found anywhere in cognitive research. This is what makes the face the choice paradigm for illustration of problems in multidimensional representativeness. Varying only as many variables as factorial design can handle implies, as we have seen, either the nullification of the others, which in turn leads to disengagement of the physiognomic response machinery, or their being held constant at a finite value, which in turn leads to fixation of impression. Taking up a few factors at a time obliterates a good part of configurational interaction. Only representative design can bring out the relative strength of factors as well as their interaction equitably; at the same time it can do so economically by confining itself to existing configurations. The situation is the same for all high-complexity problems of life and behavior.

A good access to the utilization problem is given by a study of the practices of professional make-up men. Much relevant information can be found in such compendiums as that by Strenkovsky (1937).

6. STUDIES PICTURING THE FULL HUMAN FIGURE: IMPRESSION AND ECOLOGICAL VALIDITY

As was pointed out earlier in this book (§ VIII, p. 49) the only physiognomic cue with well established, even though low, ecological validity relative to tested intelligence is height (and perhaps also weight or width) of body build. Since the functional approach to perception demands that ecological validity and utilization be viewed together and studied as to possible parallelism, the inclusion of the full stature in studies of physiognomic impression may be considered to add a representative note to the choice of experimental variables in the design. The following two experiments have considered this factor.

The first is by this writer (1939c) and is still relatively close to the systematic approach represented by the schematized-face experiment. It even uses two of its faces as part of its stimulus material. The design and results are summarized in figure 40. The so-called "medium face" (*m*) is identical with the Norm H5 in figure 36 (also shown in fig. 35); *m* or proportionate distortions of *m* are used in all but one of the eight schemas shown in figure 40. The unfavorable face (*u*), used for schema M_u only, is the impression-extreme T7 from figure 39. The "medium build," arrived at after some trial and error so as to obtain the impression of normal proportions, is called *M*. The combination of *m* with *M*, M_m, furnished the Norm which serves as a pivot for all variation (and scores for which are therefore repeated for each comparison in fig. 40). Simple proportionate magnification and reduction of M_m yields the large (*L*) and the small schema (*S*), and uniform lateral distortion yields the broad (*B*) and the narrow (*N*). B_m and N_m have the build of *B* and *M* but an undistorted face.

Preference indices, obtained by a ranking procedure somewhat related to that used for the schematized faces, reveal that there is a general tendency to favor the medium and all enlarged figures. Within this trend, our group of 46 men seems more impressed by broad build (B_m) and women more by height (*L*).

Except in the case of apparent energy the face seems to have considerable influence. For apparent happiness and apparent intelligence, significant differences can be found for pairs differing only in stature and not in facial proportion. Since this is not a representative experiment, no ecologically generalizable comparison of the relative influence of face vs. stature can be made, however. Addition of spectacles to the standard face (not shown) increases apparent intelligence and decreases good-lookingness.

Somewhat greater representativeness may be claimed by an experiment by Wallace (1947) in which photographs of four men differing considerably in appearance but identically clothed (by courtesy of a clothing store) were photographically distorted in a manner similar to that of the preceding

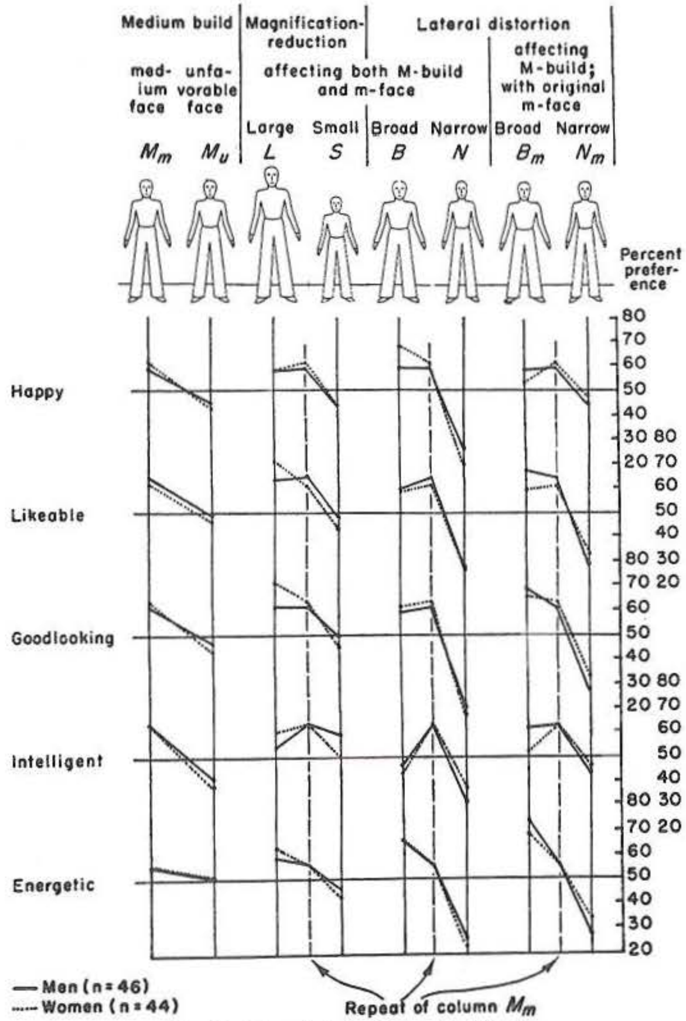


Fig. 40. (From data by Brunswik, 1939c.)
Influence of Face and Build upon the Impression Value of Schematized Figures.

experiment. For each of these four social objects or "models" there were four "versions," *M*, *L*, *S*, and *N* (in terms of the preceding experiment, with 10% increment or decrement in one or both dimensions). Further versions and models (among them one woman) were used as intermediary comparisons so as to avoid direct comparison between the four models in their different versions. A total of 46 pairs were judged in paired comparison in a setting suggestive of equal distance of the objects. Seventy students served as subjects (judges).

In table 7 the variability of the responses from model to model (with the four versions lumped together for each model) is compared with the variability from version to version (the four models lumped together within each

TABLE 7

Social-Perceptual Impressions from Distorted Photographs of Adults in Standardized Clothing

Impression quality	Variability among four main *models* (using mean of four versions for each model) S.D.	Variability among four main *versions* (using mean of four models for each version) S.D.	Variance ratio for models (persons) vs. versions *F*
Likability	10.9	2.2	24.6[a]
Happiness	13.1	2.6	25.4[a]
Good-lookingness	12.3	3.7	11.0
Energy	15.6	2.2	50.3[b]
Intelligence	14.5	1.7	72.7[b]

Source: After Wallace (1947); last column added.
[a] Significant between .05 and .01 levels.
[b] Significant at .01 level or better.

version). The variance ratios show that except for good looks the former is significantly larger than the latter; significance is established in spite of the small number of models and versions. Gross build (to the extent varied here) thus seems less influential than face and/or the details of build (or bearing) as they characterize the individual. In addition, size alone was also a statistically significant determiner of the ratings, with the influence of height strongest, proportional magnification next, and widening least (not shown in table 7). This trend of the results is in line with that of the preceding experiment with distorted line drawings.

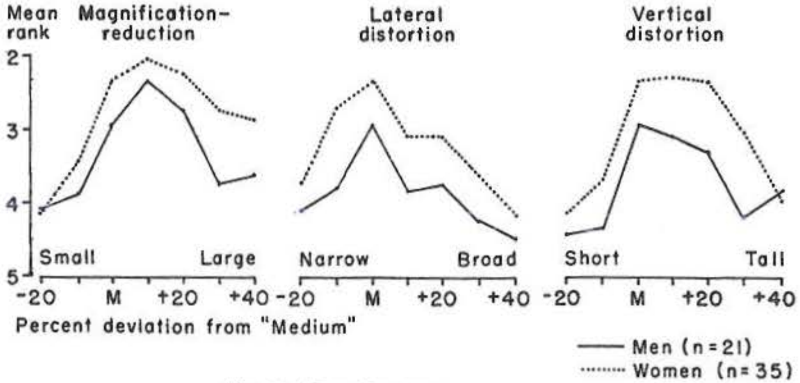


Fig. 41. (From data by Wallace, 1947.)
Apparent Intelligence in Full Photographs of Actual Persons Moderately and Extremely Distorted.

One of the non-representative features of the experiment just described is the confinement of distortion to a range which, although wide, may in consequence of a possible depth effect of size perhaps not have been quite large enough to include extremes of stature. Wallace therefore undertook a sequel to his experiment, using distortions from −20 per cent to +40 per cent. In this additional series only one fairly average model was used in nineteen versions. These were interspersed, one at a time, with the medium versions of the other models arranged in a row; ranking (from 1 to 6) was used to determine changes in preference from one version to another. As is shown in figure 41, the ratings, by 56 judges, tend to become unfavorable under the "medium" and above the +20 per cent versions.

What has been said above in discussing compromise designs for the study of facial impression also applies here: limitations in ecological generalizability in the technical sense persist until we decide for a clean shift to representative design.

XVI. Textural Ecology as a Propaedeutic to Functional Psychology

From the systematic-representative hybrid designs which have occupied our attention in the preceding chapters we turn once more to fully representative designs. However, in contrast with the fully representative studies in social perception (§ VI) and in size constancy (§ VII/3) which involve both ecological and functional validities and therefore must employ at least one subject, the problems discussed in the present chapter are concerned with ecological validity only. Since they deal with interrelationships among external variables which have to do with the causal texture of the environment they may be labeled studies in textural ecology. Since the ignoring of organismic responses, if any, should prevent even the thought of a responder-populational generalization of the results, textural ecology brings home rather dramatically the basic purpose of representative design, which is ecological generalization.

Thus far we have encountered problems of ecological validity mainly in the analysis of the objective trustworthiness of physiognomic cues (§ VI, esp. fig. 6, left; § VIII/1; and § XV) and of size and depth cues (§ VII/3, esp. fig. 8; and § VIII/2, including fig. 10, left). Here we will begin with a survey of a recent study touching upon Gestalt principles and upon the controversy between Gestalt and learning theory. Next we will turn to depth cues again.

1. Ecological Validity of "Nearness" (Proximity) Relative to the Mechanical Coherence of Manipulables

One of the best known of the organizational principles of human perception listed by Wertheimer (1923) is the factor of "nearness," also called "proximity." According to this principle, the two constituent items in pairs of dots or parallel lines will the more tend to be united in a common perceptual "figure" the closer to each other they are in the visual field. The same principle was also found to hold in animals. For birds it was experimentally confirmed by Hertz (1928) and for rats, under certain conditions, by Krechevsky (1938). A case of the utilization of the proximity principle in music has been discussed by Arnheim (1954, p. 57).

The functionalist must ask whether or not organization by nearness is biologically useful. Such usefulness would be given if, more often than not, parallels close to each other in the visual field could be traced to the common boundaries of a mechanically coherent, manipulable object; while, on the other hand, parallels far apart would reveal themselves more often to delimit mechanically irrelevant incidental arrangements, such as interspaces (interstices), ornaments or shadow areas. Clearly, the problem is one of ecological validity; in this case mechanical coherence vs. non-coherence is the distal, and nearness (or proximity) in the field the proximal variable.

The problem of the ecological validity of nearness (or proximity) can be answered only by representatively sampling the existing ecology. This writer and Kamiya (1953) attempted to approximate such a sample with the use of seven shots from a well-known motion picture reproduced in a popular magazine and covering a well-distributed variety of common-life situations (fig. 42). Only pairs of adjacent straight (or near-straight) parallels (or near-parallels, up to 5° deviation) were considered; pairs separated by a third parallel were omitted. A total of N = 892 pairs were discernible in the original pictures. From the proximal point of view, their separations were classified in eight distance categories ranging from "smaller than ½ mm" through "½–1 mm," "1–2 mm," and so forth in geometric progression, to "32–64 mm." From the distal point of view the pairs were classified dichotomously (A) as forming the joint contour(s) of a bilaterally exposed relatively permanent mechanical unit or (B) as falling into any one of a group of other alternatives having in common a more incidental and behaviorally less relevant origin of the separation.

A list of the distal classifications follows; the number of line pairs in each category, the approximate geometric mean of the corresponding proximal separations, and examples from figure 42 are also given.

(A) *Mechanical units*, fully exposed (N = 334; Geom. Mean = 1.2 mm). This category comprises all separations in which both lines of the pair represent mechanical boundaries of one and the same mechanically coherent object. Examples are, in I, the width of the crown of the hat worn by the man in the center (a), or, in VII, the vertical tree trunk in the background between the two men (q).

(B) *Relatively incidental and/or mechanically irrelevant separations*. These fall into the following subgroups:

Interspaces (N = 171; Geom. Mean = 2.8 mm). Under this category are subsumed interstices, passages, and overlapping recesses, that is, separations representing holes, gaps, or spaces between mechanical units; these separations traverse a depth or recess, the background fill being intercepted on both sides by other objects. Examples are, in I, the dark doorway (b), or, in VI, the curved dark background area between arch and lamp (o).

Ornaments (N = 204; Geom. Mean = 1.3 mm). This category comprises ornamental divisions, that is, separations defined by distinct and regular color inhomogeneities on a

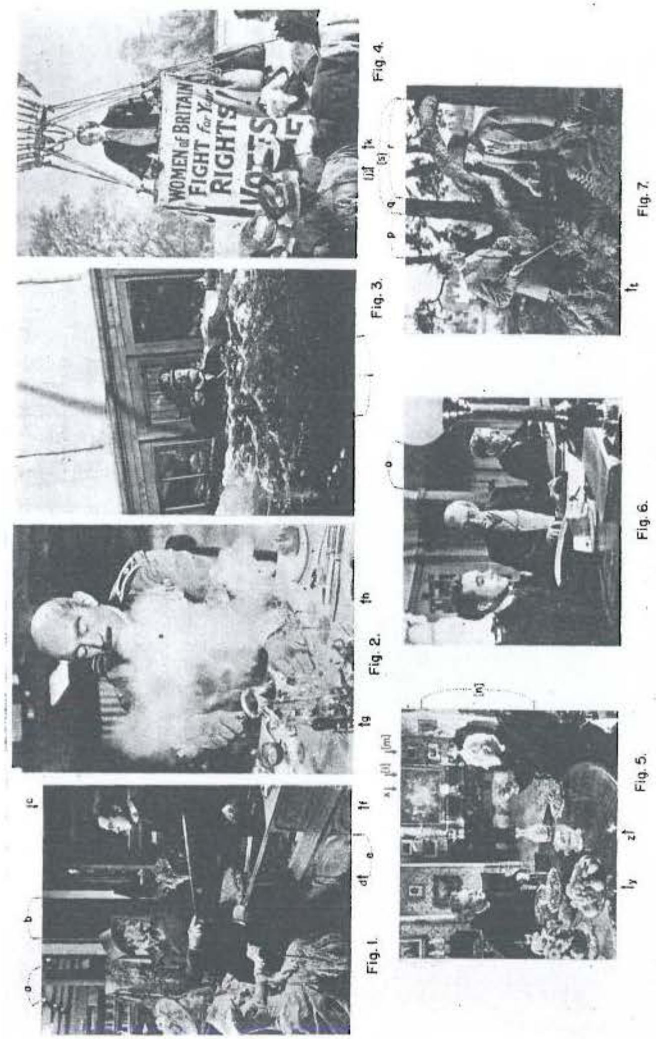


Fig. 42. (After Brunswik and Kamiya, 1953.)
Representative Design on the Ecological Validity of the Gestalt Factor of "Nearness."

coherent surface, such as stripes or flat moldings, so long as they are not caused by light and shade distribution. Examples are, in II, the vertical stripes on the object behind the officer's right upper arm (g), or, in IV, some of the letters on the large sign in the center (k).

Overlaps (N = 56; Geom. Mean = 1.7 mm). These are one-sidedly covered mechanical units or separations in which one of the lines represents the mechanical boundary of one object, the other the mechanical boundary of another object overlapping the first. Examples of such protruding object parts are, in VII, the bright lower part of the coat of the man at left extending beneath the right coat sleeve (t), or, in I, the visible end of the white cuff on the clerk's left wrist (f).

TABLE 8

ECOLOGICAL VALIDITY OF GESTALT FACTOR OF NEARNESS (PROXIMITY) RELATIVE TO MECHANICAL OBJECT COHERENCE

Distal variable "Mechanical coherence" (dichotomous) +	−	Proximal variable "Nearness" (continuous) + −	N	Point-biserial r between distal and proximal variable	CR
Mechanical units vs........	Interspaces	Small vs. large separation between parallel lines in the picture	505	+.34	7.2
	Interspaces and all others		892	+.12	4.0

SOURCE: Adapted from Brunswik and Kamiya (1953).

Shadow-bounded separations (N = 33; Geom. Mean = 2.8 mm). This category defines separations for which one line of the pair is a shadow contour and the other represents a mechanical boundary of the object upon which the shadow is cast. Many items in this category do not represent shaded portions of the field but normally illuminated areas "left over" after deduction of a shadow. Examples are, in III, the lighted expanse of the upper end of the ship's door above the head of the captain (i), or in I, the separation between the edge of the panel in the lower right corner of the picture and the right contour of the umbrella shadow mentioned below (e).

Shadows (N = 94; Geom. Mean = .7 mm). As defined for our purpose, this category is limited to separations between two shadow contours (penumbras), or between one shadow contour and a mechanical boundary of the object casting the shadow. An example of the former, bilateral type of shadow is, in I, the shadow cast by the umbrella handle on the panel of the cabinet in the front part of the scene (d); an example of the latter, unilateral type of shadow is, in II, the shadow of the knife blade on the table in the lower right-hand corner of the picture (h).

The ecological validity of nearness relative to mechanical coherence can already be gathered from an inspection of the geometric means. The mean for (A) is only 1.2 mm while it is 2.8 for interspaces and stays above the mean of (A) for all the other subcategories of (B) except shadows, a relatively unpopulous subgroup.

A better way of appraising the ecological validity is by the usual measure, the correlation coefficient. As shown in table 8, the validity is .34 when only interspaces are considered under (B); it drops to .12 when the entire material is included. As can be seen from the critical ratios between means (customarily used with the point-biserial r) both coefficients are statistically highly significant, however, and thus ecologically generalizable within the limits prescribed by statistics.

The lowering of the correlation when the entire material is included is due to a large extent to those narrow "shadows" which run along, and help to reinforce, object contours. Since strong artificial lighting was used, our material is perhaps not quite representative with respect to these shadows. Be this as it may, we have singled out this subgroup because of its peculiar behavior. The correlation with (A) becomes negative for this last subgroup (–.22; CR = 5.1).

In order to eliminate the possibility that a narrow picture separation be caused by a wide separation in a tilted object, those 483 of the 892 line pairs for which both constituents appeared to issue from (approximately) the same depth were singled out and the analysis repeated. The results reported above were confirmed; all coefficients were found either slightly higher (.37, and .13, respectively), or less negative (–.19).

2. AUTOCHTHONOUS GESTALT DYNAMICS VS. PROBABILITY LEARNING OF CUES

According to orthodox Gestalt theory, perceptual organization factors operate by virtue of the field-dynamics inherent in the brain processes rather than by some functional principle based on experience; the possible ecological validity of the respective stimulus configurations is overlooked or even denied (Köhler, 1947, pp. 107, 132 f., 142; Koffka, 1935, pp. 160, 164–167; see also Brunswik and Kamiya, 1953, pp. 20–23).

The proof, just reported, of the existence of such a validity at least for one of the Gestalt factors of organization opens up the possibility of viewing the stimulus configuration involved in this factor, proximity, as a cue acquired by generalized probability learning. In the same study we have found preliminary evidence of the ecological validity of other Gestalt factors—symmetry and closedness—also; in fact, validity coefficients promise to be much higher values for these than for nearness. Thus it may well be that all the Gestalt factors may be open to reinterpretation as externally imposed upon, rather than as innately intrinsic to, the processes in the brain; they all would then appear as functionally useful rather than as whimsically "autochthonous."

It goes without saying, however, that such a functionalistic reinterpretation of allegedly "dynamic" principles would lose much of its cogency if it would turn out that the configurations in question possess similar organizing effects in individuals, groups, or species in whose habitat or culture they have no (or negative) ecological validity.

Furthermore, the successful demonstration of the ecological validity of a Gestalt factor does not automatically imply the legitimacy of its interpretation as a learned cue. It merely shows that an objective basis for probability learning is offered the individual within his surroundings. Since, however, all ecological validities represent a challenge to the organism for utilization, and since it appears that certain cues are on the average being utilized

roughly in proportion to the degree of their validity (see above, pp. 43, 50 ff.; see also § XV/6) our findings lend plausibility-support to a viewing of the Gestalt factors as cases of successful cue-utilization subsumable under the principles of learning theory.

3. MORE ON THE ECOLOGICAL VALIDITY OF DEPTH CRITERIA

It will be remembered that in figure 10 a few question marks were inserted to indicate the tentativeness of Seidner's results concerning the ecological validity of some of the well-known perceptual depth criteria. In the meantime, Seidner continued his ecological analysis of these criteria with a set of 30 original photographs of scenes from the daily routine of a specific person. Subsequently, on-the-spot depth measurements were taken for 20 points per picture, the points being selected with the use of random-number tables. A total of 600 points (300 point pairs) was thus available for analysis; for a sample of this size, correlations of .08 and over are statistically significant.

Coefficients obtained for the three depth cues discussed on page 49 f. are being presented here, again with the kind permission of Mr. Seidner. They are in part lower and in part higher than those reported on the basis of the original sample of 75 point pairs from magazine pictures; for each cue a reasonable definition can be found under which significant validities are obtained. Problems of definition center about such questions as whether or not sky-points in the picture should be included (assuming a fixed, rather large distance for them), or whether extreme colors or a neutral gray be taken as the zero point in defining the color cue.

The validity of "vertical position" as an indicator of real distance was tentatively given as .61 in figure 10. With the new sample it was found to be .41 with sky-points included, .21 with sky-points excluded; both these values are significant.

The validity of "space filling" was given as .42 in figure 10. Now it is about .10—still significant—with sky-points included; it is .15 or .22 without sky-points, depending on whether the total number of discernible inhomogeneities between the two sample points or their density per length unit on the picture is considered. In this context it must be kept in mind that loss of detail in the photographing of situations at greater distances tends to obscure the validity of this cue.

The validity of "color" (albedo—"brightness"—in the picture) as a cue for distance was given as .23 in figure 10. Now it is as high as .39 with sky-points included; but it becomes insignificantly negative (–.06) with sky-points excluded. The values remain very similar (.36 and –.05, respectively) when extreme albedos are taken as negative and a medium gray as positive.

XVII. ACQUISITION AND EXTINCTION OF PERCEPTUAL CUES

Since ecological validities may change from one habitat to another, a certain flexibility of the perceptual system in the utilization of cues would seem to be adjustive. We briefly turn to this problem now.

In recent years evidence has multiplied that the so-called primary depth cues, such as binocular disparity, may function without prior individual experience and that they may be relatively resistant to modification by experience (Hess, 1954). The secondary or "empirical" cues, on the other hand, are undoubtedly more flexible. In an earlier study Hess (1950) has presented results on the development of the light-and-shade cue in chickens which point to a combination of innate and acquired influences with a predominance of the latter.

1. ACQUISITION OF DEPTH CUES INCIDENTAL TO SYSTEMATIC EXPERIMENTS

The fact that new depth cues may be acquired and extinguished rather casually and in an *ad hoc* manner is suggested by some of the data on size constancy by Holaday (1933, pp. 454, 462) and on shape constancy by Eissler (1933, pp. 504, 529). Both authors employed a viewing of the experimental scene through a photographic camera in some of their series. Constancy ratios are lowest under these conditions. For size constancy the average c is only .03, and several of the individual constancy ratios are negative; this is considerably below the most unfavorable figures reported above in table 1 (p. 25), that is, those for dark adaptation and darkroom. For shape constancy even the average c is negative, –.12. On the other hand, at least size constancy is known to be fairly high in viewing ordinary photographs (Gibson, 1947, p. 212).

In an effort to explain their surprising results Holaday and Eissler point to the fact that in their camera series Standard and Comparison had exchanged positions on the screen. Since in all the other series of the study the lateral position of the relatively near and the relatively far object remained the same, the conclusion seems inescapable that lateral position had become a depth cue in the framework of the experiment.

This conclusion is confirmed by direct observations of Eissler in certain control series in which the Standard was presented in a frontal plane rather than under an angle. A slight phenomenal tilt was seen, usually amounting to no more than 2 to 3 degrees, while actual rotation of the Standard had ranged from 30 to 75 degrees; in addition, this perseveration effect diminished rapidly and thus the new cues appeared to be readily extinguishable.

2. ARTIFICIAL CUES OF ILLUMINATION IN COLOR CONSTANCY

The problem was followed up in an experiment by Fieandt (1936). One of the familiar concealed-shadow experiments (see Woodworth and Schlosberg, 1954, pp. 441 ff.), in a version by Kardos (1934), served as the starting point. Figure 43, A, shows a light gray aperture board with a large circle cut out. The board is placed some distance from the back wall of a not too well lighted room. A smaller disk of the same light gray paper is suspended in the center by means of invisible threads attached to the back of the board. The board receives additional illumination from a light-source placed somewhat to the right; a shadow-caster shaped like the shadow in C but smaller is introduced between board and light-source in such a way that only the disk is affected. Since under these circumstances the shadow penumbra falls within the ring-shaped aperture between disk and board and thus is not in evidence, there is no color- (whiteness-) constancy and the disk appears as if cut from a paper of different, dark-gray albedo. When the shadow-caster is moved closer to the light-source the shadow increases as in C; its penumbra which is now on the board establishes color constancy both for the shaded inner rim of the board and for the disk, so that the latter whitens up dramatically in appearance. The fact that this color constancy effect disappears upon return to stage A and the disappearance of the specific penumbra cue

may be taken as one of the evidences for the functional autonomy of perception (see § XIV/3).

It may be added that by suitable arrangement the shadow cast by the entire experimental set-up on the background wall to the left of the aperture board can be brought in plain view of the observer. The fact that a shadow is cast on the center disk is given away by this image in a rather tell-tale fashion. Yet perception does not respond as one may expect of a sensible being; it persists in responding only to the local cue, the penumbra. Here

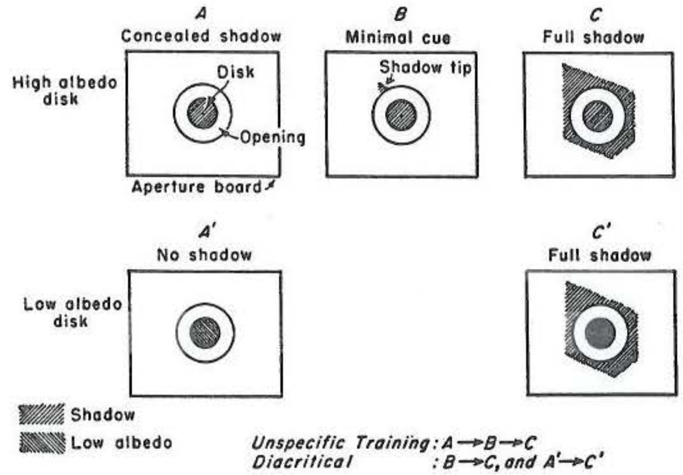


Fig. 43. (To a study by Fieandt, 1936.)
Schema of Concealed Shadow Experiment in Perceptual Learning.

the autonomy of perception is not only relative to intellectual information but also to non-routine cues coming from within the subsystem of perception itself. Stereotypy and lack of integration, or what the writer (1934) has called the inherent "stupidity" of perception, are well brought out by this example.

Fieandt's modification of the original Kardos experiment consisted in repeating the changeover from A to C a large number of times and in stopping at the intermediate stage B in order to ascertain whether or not the incipient triangular shadow tip shown in the picture could be conditioned as a cue for the true illumination of the disk. As shown in figure 44, left (averages of two subjects), the disk in B brightens up considerably in the course of 100 trials, as checked every 20 trials by means of an adjustable rotating Comparison disk. Even A participates in the effect (while the slight rise in C must be interpreted as a relatively incidental control effect).

To divorce the training from the total situation and make it more specific relative to the shadow tip, a diacritical experiment was undertaken next. In situation A' there was no concealed shadow but the center disk was replaced by an actually more grayish disk (one of a paper of lower albedo), carefully selected so as to simulate situation A in its retinal impact. Upon introducing the shadow-caster as in C′ it became evident to the observer that the disk was in reality dark gray. In this experiment stage A was omitted from the training sequences (even though not from testing), and transitions B–C were interspersed with sequences A′–C′. Figure 44, right, shows (averages of eight subjects) that the constancy effect for B in-

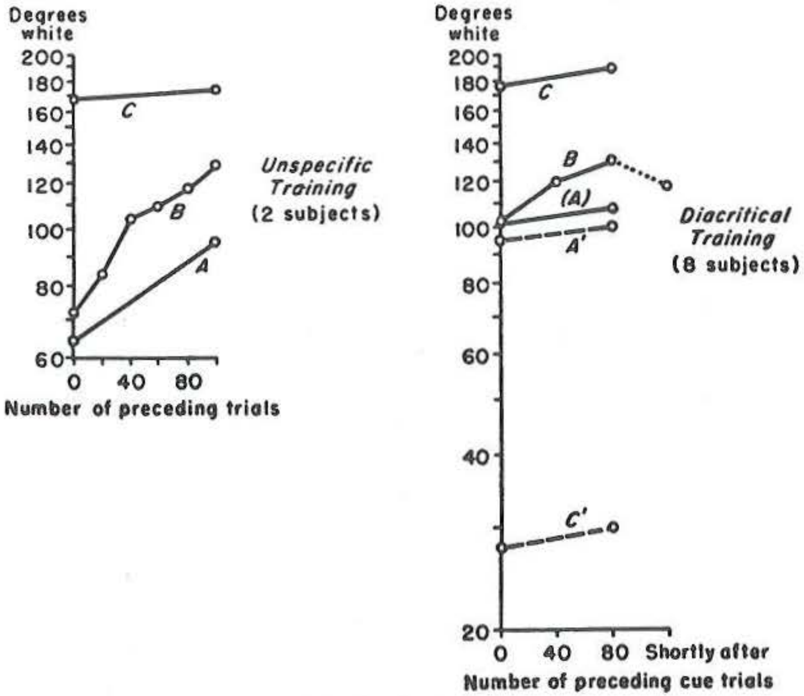


Fig. 44. (From data by Fieandt, 1936.)
Perceptual Learning of Cues of Illumination.

deed reoccurs over the 80 trials now employed, but that A now remains unaffected (except for joining all the controls in their slight incidental rise). The shadow tip as a minimal cue thus had been acquired by the perceptual system as a diacritical illumination cue separating the color of the disk in A from that in A′.

The transient nature of this acquisition is shown by the dotted end of curve B which connects with the results taken in an after-test given soon after completion of the experiment during the same session.

Additional experiments (with a few subjects only) indicate tentatively that a small button on the disk, or even the ringing of a bell, may be conditioned as illumination cues in a manner similar to the shadow tip.

Fieandt has also found that while, as we have seen, perceptual learning

is slow and almost animal-like in its insistence on "stamping-in," it also is remarkably independent of intellectual insight into the specific cue acquired, and vice versa.

In a subsequent study (1938) Fieandt has expanded and confirmed his results as reported here.

3. CUES OF LIMITED VALIDITY: PARADOXICAL DECLINE OF THE PROBABILITY LEARNING CURVE AND A NEGATIVE RECENCY EFFECT

An added representative feature in experiments on cue learning and on learning in general is introduced by shifting the artificial ecological relationship from an absolute right-wrong alternative to a probabilistic or "partial" type of reinforcement. For perceptual cue learning this was done in an experiment by Brunswik and Herma (1951; preliminary report by Brunswik, 1938) briefly referred to on page 56. Sequences of weight pairs were designed to present heavy weights more often (⅔ of the time) on one side, drastically lighter weights correspondingly more often on the other side for simultaneous lifting. The artificial-random sequence of combinations for the two sides is shown at the top of figure 45; of each unit of nine presentation pairs four are confirming of the trend on both sides (+), one is infirming on both sides (−), and four are confirming-infirming, or "neutral" (0). The series of cumulative ecological validities of position as a cue for weight-to-come which is the result of this design is shown in the upper curve of figure 45. The curve is presumed to start at zero; the first few trials create considerable fluctuation, and later the validity curve settles in decreasing pendulum swings about $\phi = .33$ (corresponding to 67 per cent).

Since in the neutral trials either both weights are heavy or both light, these trials were used to check on the possible development of an "expectancy" illusion or successive weight contrast by asking for simultaneous comparison. Results are shown in the lower curve of figure 45. The dotted vertical guide lines connect the test trials in the presentation sequence at the top with the corresponding per cent frequencies of illusion in the lower curve, passing between adjacent points of the cumulative ecological correlation curve. The rise of the learning curve—based on 36 subjects—shows that an expectancy illusion is indeed created in response to the artificially established position cue. It reaches its maximum at the 11th trial; then it tapers off in a "paradoxical" decline despite continued positive reinforcement of the over-all ecological validity.

This paradoxical decline may be either a "disappointment" feature specific to prolonged probability learning, or the artifact of our use of an illusion as the test of the learning of the cue. The former interpretation has found tentative support in a study by Jarvik (1951). This study concerns a type of expectancy that is more intellectual than perceptual. Three randomized series, A, B, and C, in which one word, "check," held a preponderance over another, "plus," at the rate of 60, 67, and 75 per cent, respectively, were presented to different groups of college students totalling 78. These had to guess, before each of the 87 presentations, which of the two alternatives

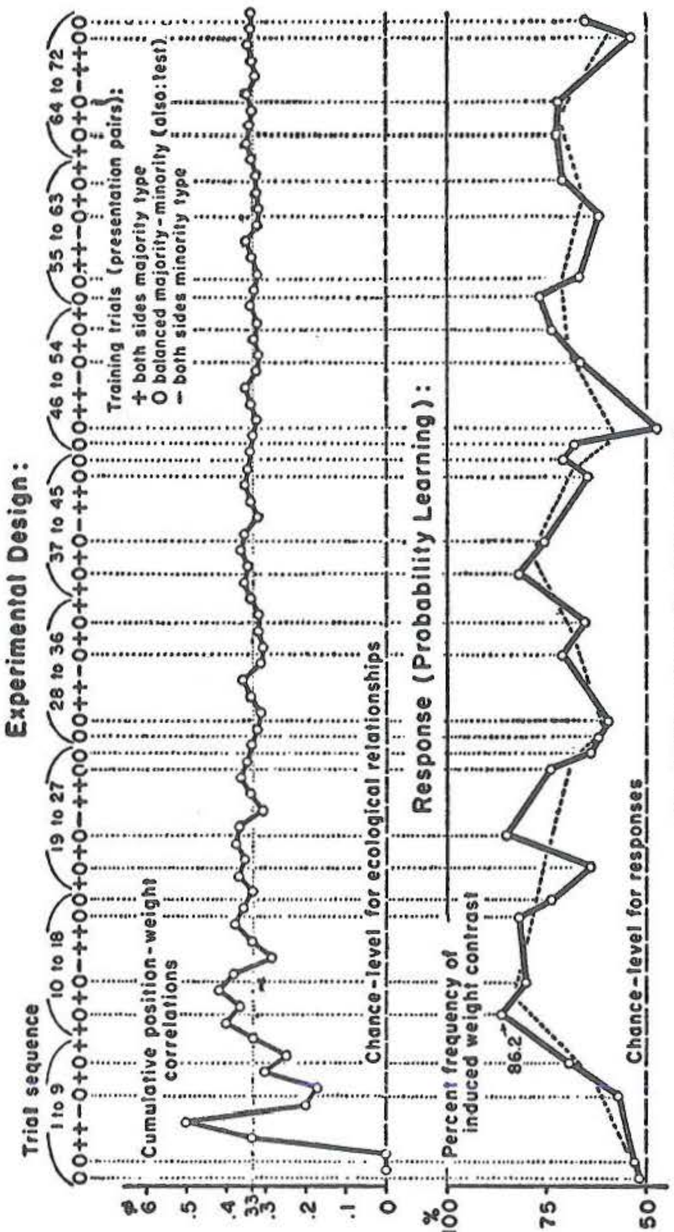
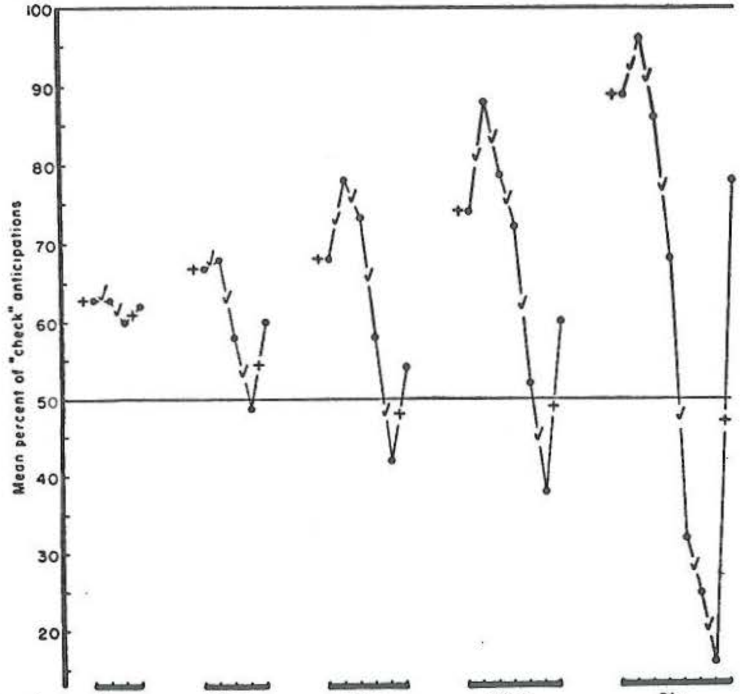


Fig. 45. (From Brunswik and Herma, 1951.)
Perceptual Probability Learning of a Position Cue for Weight.

would be next. The trend of results again shows an initial rise followed by a slow "paradoxical" decline, without reaching statistical significance in this latter respect, however.

Superimposed upon, and occasionally overshadowing, the long-range positive probability learning Jarvik found a short-range "negative" recency effect; this consisted in an increased anticipation of the opposite alternative in cases of serial accumulation of one of the alternatives in an incidental "run."



Length of runs		Two "checks"	Three "checks"	Four "checks"	Five "checks"	Six "checks"
Number of runs	A	10	4	1 (*72-77)	–	–
	B	6	6	4	1 (*69-75)	–
	C	7	3	6	1 (*76-82)	1 (*51-58)
Total runs		23	13	11	2	1

Fig. 46. (From Jarvik, 1951.)
Negative Recency Effect for Prolonged Runs in Probability Reinforcement.

In figure 46 average results of consecutive runs ranging from two to six "checks" in a row are presented. Note first that as we move from the left-hand to the right-hand graphs in the figure, curves start higher and higher. This is a direct result of the fact that extended runs of "check" reinforcements are more frequent in series with higher over-all reinforcement ratios and thus with a more pronounced learning effect. Also, the longest runs happened by chance to be in the later portions of the training (as can be seen from the bottom part of the figure).

After the first "check" reinforcement there usually is a tendency toward a rise in the curves, in conformity with the classical positive recency effect.

The negative recency effect sets in after the second "check" reinforcement; the curves drop consistently despite continuing reinforcement. In other words, there is an increasing tendency on the part of the subjects to expect a continuous sequence of the generally more frequent of the two alternative words to lead to a change-over to the other alternative on the next trial. The interference of this negative recency effect with the over-all trend of probability learning is so great that after three to four "checks" the gains are as good as obliterated, and after four or five "checks" the chance-level is crossed and the preponderance of anticipations is now in the opposite direction, that is, of "pluses" to come. In the single long run of six consecutive "check" reinforcements in the second half of Series C as few as 18 per cent anticipate a further "check" to follow. The "plus" presentation that terminates each run brings "check" expectancies back above chance even though not quite to where they were before the run. Similar results were obtained for extended "plus" runs.

Both paradoxical decline and negative recency effect are phenomena specific to probability learning. Just as perceptual size constancy cannot be studied in a tied-variable design of classical psychophysics because of the impoverishment of the ecology, the learning phenomena mentioned are by their very nature unamenable to study under the conventional black-white (100 to zero) reinforcement policy. They can emerge only with the introduction of probabilitistic policies more representative of reality.

More specifically, Jarvik's discovery of the negative recency effect owes much to the fact that in his case representativeness of design was carried still further. Instead of the customary so-called artificial random series he used real randomness. His sequences had been determined by the use of Fisher and Yates' random-number tables within the predetermined relative frequencies; this policy led to an at least occasional occurrence of the excessively long runs of the same symbol which brought out the negative recency effect so drastically that it could no longer be ignored.

4. PERCEPTUAL LEARNING *Against* AWARENESS OF WHAT IS LEARNED

As a further contribution to the problem of autonomy of the perceptual system we may add that the subjects' awareness of the cues or frequencies was as a rule dim in both of the studies reported in the preceding subsection 3; if present, it was frequently grossly erroneous and seemed to have little to do with performance.

The fact that in the Brunswik and Herma experiment only the test trials were used for verbalized judgment, together with the prevalence of the induced contrast illusion in these trials, made for a reversal of the weight-judgment and the weight-reality trends; when heavy weights were more frequent on the left, "heavy" verbalizations tended to be more frequent on the right, and vice versa. The question arose as to whether awareness would follow the real trend or the verbalization trend.

Twenty-two of the subjects were asked after completion of the experiment: "In your opinion, which objects were on the average heavier, the right or the left?" Those subjects who seemed uncertain as to whether they should refer to the judgments actually asked for, or to their impressions in the total trials regardless of whether or not these impressions were overtly expressed, were told to refer to the latter.

Conscious impressions of weight trends in terms of right vs. left position turned out to be *in opposition* to the reality trend in four-fifths of the cases. In our case at least, explicity conceptualized learning thus is determined primarily by overt verbalization trends rather than by the objectively—and undoubtedly also subjectively—far more drastic opposite differences in the more crucial part of the training. Yet the emergence of our perceptual contrast effect reveals the simultaneous acquisition, at a more primordial level, of a concurrent implicit quasi-expectancy in the correct direction. This is not only a case of "submerged mediation" (Brunswik, 1934), or a case of what since Thorndike has become known as "learning without awareness of what is being learned;" it is learning *against* awareness of what is being learned. Restriction of verbalization to the test trials has made it possible for the two levels of learning, the perceptual and the verbal, to stand out more clearly against each other

These results, together with those reported in subsection 2 from the study of Fieandt, underscore the unconscious component in perception. Along with such features of perceptual functioning as stereotypy, slowness of learning, and so forth, which we have emphasized in various contexts earlier in this presentation, they help to consolidate the image of perception as a primitive "id"-like sub-system of cognition (Brunswik, 1934, pp. 118, 127 ff., 224). This primitivity must be held against such compensating virtues of the perceptual system as the rapidity of its response and the relative mellowness of its error distribution which we have emphasized in comparing it with its higher-order rival, thinking (§ XIV).

XVIII. Problems of Stimulus Representativeness in Clinical Psychology[23]

1. PERSONS AS STIMULI IN SOCIAL AND CLINICAL SITUATIONS

The most dramatic case we could make for representative design was that of the perception of social objects (§ VI). In this case both subjects and objects are persons; if the former are to be sampled representatively this could hardly be denied to the latter. Yet, as we have seen (p. 39 f.), the traditional ways of handling stimuli systematically rather than representatively are so ingrained that even this simple logic has prevailed only in exceptional instances.

Kenneth R. Hammond (1948) has expanded the concept of social object to include examiners and testers in social-psychological surveys and in clinical situations. Since in any testing procedure the examiner constitutes part of the external stimulus situation, representative design demands that examiners should also be sampled. In the article cited and in a second paper (1954) Hammond discusses a series of examples from the recent literature in which glaring overgeneralizations concerning the influence of sex or race of the interviewer or tester were made on the basis of an inade-

[23] This chapter is based on a paper presented as part of the *Symposium on the Theoretical Basis of the Rorschach* held at the Veterans Administration Research Conference in Clinical Psychology, Oakland, California, April, 1953.

quate selection technique borrowed from the traditional systematic rather than representative treatment of physical stimuli in classical experimental psychology. Hammond found the same paradoxical discrepancy between the sampling of subjects proper and the sampling of examiners which this writer had found in studies on social perception from photographs (p. 40).

In one of the studies criticized by Hammond (Curtis and Wolf, 1951) the Rorschach was applied to $n=586$ subjects, and there were $N=10$ examiners, 7 men and 3 women (using, as we have suggested, n for the size of the subject sample and N for the size of the object sample). The statistical significance of sex-of-examiner was then surreptitiously claimed with the use of 586 rather than of 10 (or of 7 versus 3) for the size of the sample. Under representative design, as we have seen (p. 38 f.), there must be a second index of significance which would tell us how far we can generalize ecologically. Even texts in psychological statistics, being interested in formal more than in material problems, are as a rule oblivious of this double jeopardy of generalization.

2. STIMULUS BIAS IN THE CONSTRUCTION OF TEST PATTERNS

From the point of view of principle, in the construction of tests physical stimulus patterns would have to be representatively sampled in the same manner as persons. For the time being one may prefer to think in terms of more modest objectives, however.

As we have seen (p. 52 f.) certain tendencies in the direction indicated have existed for some time in the field of testing. But even when some sort of sampling of tests was explicitly envisaged, the sampling was usually done from an artifact, such as the multitude of tests already in use. This in turn merely reflected the bias of the earlier testers or of the laboratory experimentalists who had furnished the basis for the existing tests.

The temporary expedient we will consider here as a practicable start out of these methodological difficulties consists in an effort to base the design and evaluation of tests on a kind of canvassing or proportional representation of the major theoretical positions that have been stated or that seem conceivable in our time. Thus we would at least remove the blindness which is inherent in any clinging to experimental tradition or fashion, *per se*, and the haphazardness which goes with reliance on the test-constructors' intuitions.

We will consider the theoretical positions exhibited in "projective" testing, exemplified here by the Rorschach (1942), along with those in two variants of Gestalt psychology, the Berlin school and the Leipzig school; we will also include a consideration of some of the tests that are directly traceable to these latter schools. Gestalt psychologists share with projective testers the view that all illusory perception is not only illusion *away from* some ideally accurate or veridical type of response but at the same time is illusion *toward* something else, that is, assimilation toward some preferred patterns of organization. In projective testing most of the assimilative power is seen as coming from the inner motivational stratum with its richness of personal figures and themes. Even though Gestalt psychologists of all shadings have been

consistently oblivious of this stratum, they have been more than astute in uncovering some more strictly cognitive organizational preferences in perception. It may be claimed that not only the motivational but also the more purely perceptual tendencies of organization are laden with diagnostic potentialities.

The Gestalt psychologists and Rorschach must also be assumed to be akin in a rather deep-seated attitudinal respect. Their efforts and those of all projective techniques have in common that they are based on the recognition and utilization of what such precursors of the modern Gestalt movements as Benussi (1904) and his friends at Graz have called "Gestalt ambiguity." What this recognition may entail in the way of intellectual or personality restructuring can best be appraised in the light of the concept of "intolerance of ambiguity" as developed by Else Frenkel-Brunswik (1949) in connection with the study of the authoritarian personality syndrome and its rigid stimulus-boundness.

The most comprehensive view of preferred perceptual organizations has been offered by Sander (1928; for a somewhat less experimentally oriented and less lucid presentation in English see Sander, 1930). Sander is the leading experimentalist of the Leipzig school of *Ganzheitspsychologie*, the major rival movement of the better known Berlin school of Wertheimer, Köhler and Koffka. Sander distinguishes a series of organizational stages in the brief genesis of any single full-blown perceptual act (or of perceptual actuality); this so-called "actualgenesis" is set in parallelism to the phylogenesis and ontogenesis of perception in an over-all theory of development. A most primordial, feeling-like stage is said to be followed by a second, "geometric-ornamental" stage and eventually by a third, "realistically meaningful" (*sinnhaft-bedeutungsvoll*) stage.

For our present purposes only the second and third stages shall be considered. The geometric-ornamental tendencies are conceived of in agreement with the center-piece of Berlin Gestalt theory, the law of *Prägnanz* or "good form"; while never pinned down with ultimate clarity, the concept of *prägnant* form has been fairly well circumscribed by its originators as a tendency toward symmetry, parallelism, rectangularity, closure, circularity, good continuation, horizontal-vertical framework and related features best summarized in terms of the idealized forms of pure non-representational geometry. Hence these tendencies are also called eidotropic tendencies by Sander.

The realistically meaningful stage, which is a more specific brainchild of the Leipzig group, is said to be characterized by pressures toward representational organization in terms of palpable objects or "things." Hence these pressures are also called ontotropic tendencies, that is, tendencies toward the "existing," or, more particularly, toward the familiar.

Sander refers to numerous experiments, utilizing such techniques as extreme tachistoscopic exposition, extreme reduction in size or brightness, fragmentation, and various other kinds of impairment of the stimulus impact or evaluation including developmental materials and evidence from the arts.

All these are to serve as means of catching the perceptual act in a primitive stage. Some of the empirical evidence presented by the Leipzig school goes considerably beyond the scope of the techniques employed by the Berlin school. One of the younger exponents of the Berlin point of view, Metzger, has included some of the Leipzig evidence in his splendidly illustrated book on laws of vision (1936) but has stressed primarily the data relating to geometric-ornamental *Prägnanz*. A total of close to twenty techniques may be distinguished which, when all results are seen in synopsis, bring out the law of *Prägnanz* in an impressive variety of ways. Only a few of them have become known and have been checked over in American literature, among them such relatively unrepresentative and relatively unsuccessful ones as memory for form. The full array of experimental techniques employed furnishes an impressive example of a canvassing of relevant media in the study of a single problem.

Remembering that the law of *Prägnanz* refers to but one of at least two major types of preferred patterns of organization in the perception of form, both the Berlin research techniques and the Rorschach inkblots appear lopsided when held against the requirements of representative stimulus design. Their onesidedness tends in diametrically opposite directions, quite in the same spirit of splendid isolation that exists between the various subvarieties within the academic Gestalt movement.

The standard stimulus fare offered by Wertheimer (1923), Gottschaldt (1926), Wulf (1922) and others of the Berlin school consists of arrays or patterns of dots, straight or angular lines, and smoothly drawn curves. All these patterns are extremely provocative of geometric-ornamental or eidotropic tendencies while giving little chance to representational meaning or even to the factor of familiarity to make itself felt.

The Rorschach inkblots, on the other hand, with their peculiar tying of bilateral symmetry with irregularity and blurredness—as well as subsequently designed inkblots and the less well-known cloud-pictures which K. Struve has used under the stimulus of William Stern (1938)—give the inside track almost exclusively to the realistically meaningful tendencies. They even prove conducive to very specific types of objects (animals, anatomy, etc.) so that the natural balance of the world view is threatened even within the domain of existing physical or social objects. With an eye again on the more general controversy concerning eidotropic vs. ontotropic tendencies with which we are here primarily concerned we may note the fact that even in the more voluminous lists and catalogues of Rorschach responses no more than a handful of abstract "squares," "triangles" or other purely geometric formations could be discovered; this goes so far that even the giving of such responses by a subject is viewed with alarm in the standard evaluation of a protocol.

One way out of this dilemma would be the use of stimulus materials posing a clearcut and balanced rivalry between geometric-ornamental and realistically meaningful possibilities of organization. In perceptual testing, rivalries have so far been set up mainly along the relatively confined dimension of

color- vs. form-dominance (for some American research along this line see Thurstone, 1944). Some of Sander's designs using series of decreasingly fragmentary line-drawings (1928) may well lend themselves for adaptation to tests of the rivalry between formalism and realism. Figure 47 shows three stages out of such a series; the first elicits rather clearcut geometric-orna-

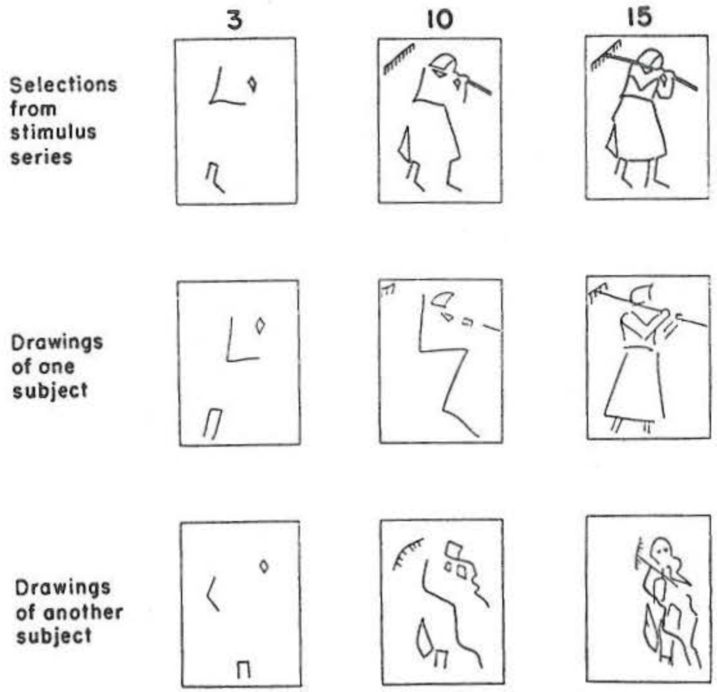


Fig. 47. (From Sander, 1928.)
Decreasing Fragmentation in Tachistoscopic Exposure and the Transition from Geometric-ornamental (Prägnant) to Realistically-meaningful Perception.

mental improvement (increased verticality, rectangularity, symmetry) while the last brings out realistically meaningful tendencies. Figure 48 shows drastic individual differences in this respect which Sander claims to be dependent on age.

Modifications of the Sander drawings were used as tests by Wartegg (1939) and by Kinget (1952). Rather impoverished stimulus configurations were employed, however, and the emphasis was shifted back from the cognitive to the emotive, projective aspects of the personality.

3. REPRESENTATIVENESS OF SCORING PROCEDURES

From representativeness as to stimulus design we now turn to representativeness as to evaluation procedures. The Bender test (1938) and its further history furnish an example of the lost opportunities that may result from one-sidedness and isolation in theory.

This test may be challenged as to representativeness in two different ways; one concerns stimulus design, and the other, evaluation. As to stimulus design (fig. 49), Bender preserves the one-sidedness of Wertheimer's classic demonstrations of Gestalt principles, and her selection is not too felicitous at that. As to evaluation, Bender has in the main relinquished whatever advantages may be offered by the stimulus materials chosen; she has largely ignored the theoretical principles which have given rise to the original Gestalt demonstrations or experiments. In the cataloguing of the responses by Bender her-

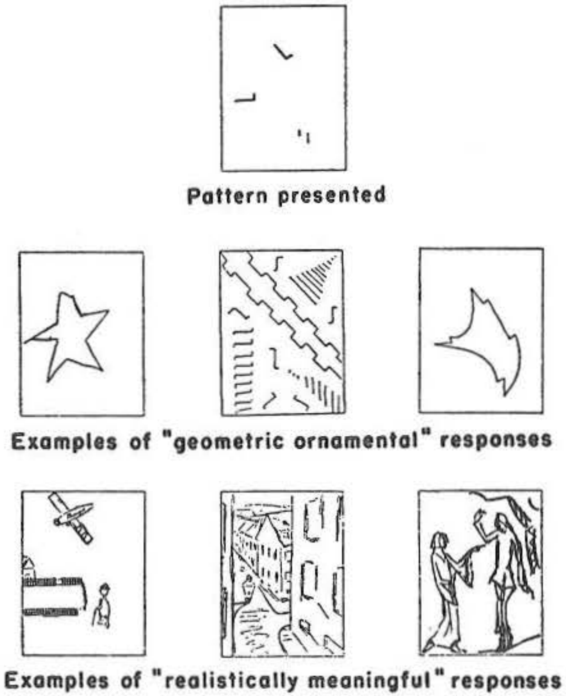


Fig. 48. (After Sander, 1928.)
From Eidotropic to Ontotropic Completion in Adolescents.

self and especially by the subsequent proponents of more rigorous scoring systems (Billingslea, 1948; Pascal and Suttell, 1951; Gobetz, 1953) there is repeated mention of such features as "circularity," "symmetry," "simplification" and so forth. Yet there is little awareness of the fact that such variables may readily lend themselves to the establishment of an over-all score and/or comprehensive subscores for *Prägnanz* which would allow us to capture the degree and kind of the geometric-ornamental formalism present in each subject. Such scores could be representative of some of the major pattern trends or organizational preferences in a given person. This lack of awareness becomes most conspicuous in Bender's discussion of her developmental findings (fig. 50). The drawings are full of confirmatory evidence for

Sander's view that geometric-ornamental tendencies are a mark of earlier developmental stages (preponderance of circularity, symmetry, horizontal-vertical framework at younger ages); yet little is made of this in the way of conceptualization or comprehensive scoring.

That the development of a Gestalt-psychologically inspired system of scoring is not merely an academic matter but rather one that may have far-

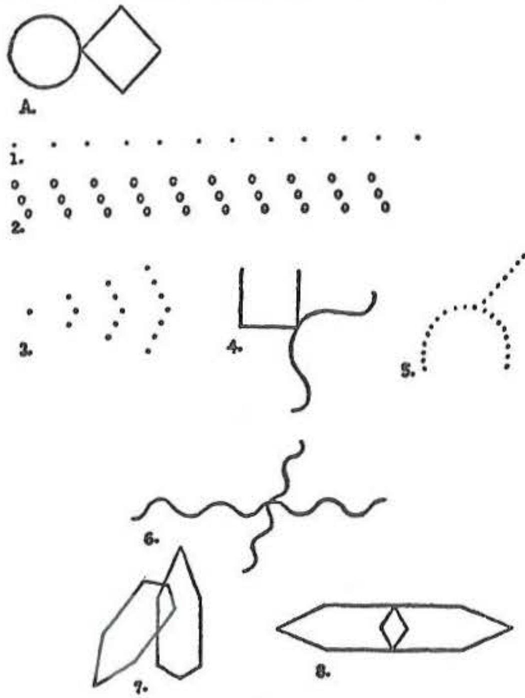


Fig. 49. (From Bender, 1938.)
Wertheimer's Figures Used in the Bender Gestalt Test.

reaching diagnostic and personality-theoretical implications is best indicated by the stress that has frequently been placed on the possible relationship between schizophrenia and a general "formalism," on the one hand, and between the manic-depressive syndrome and a general "realism," on the other. We may add to this the aspect of regression along the developmental continuum that must be inherent in formalism if there is any truth in Sander's view that the geometric-ornamental tendencies represent a relatively primitive stage. All this is reason enough to press for a formalism- or eidotropism-score (or set of scores) on the Bender test.

The customary confinement to a single over-all score on the Bender, as well as many elements in Rorschach scoring, proceed on the tacit assumption that deviations from the norm set by the stimulus are nothing but deviations *away from,* and thus must be treated primarily as losses of accuracy.

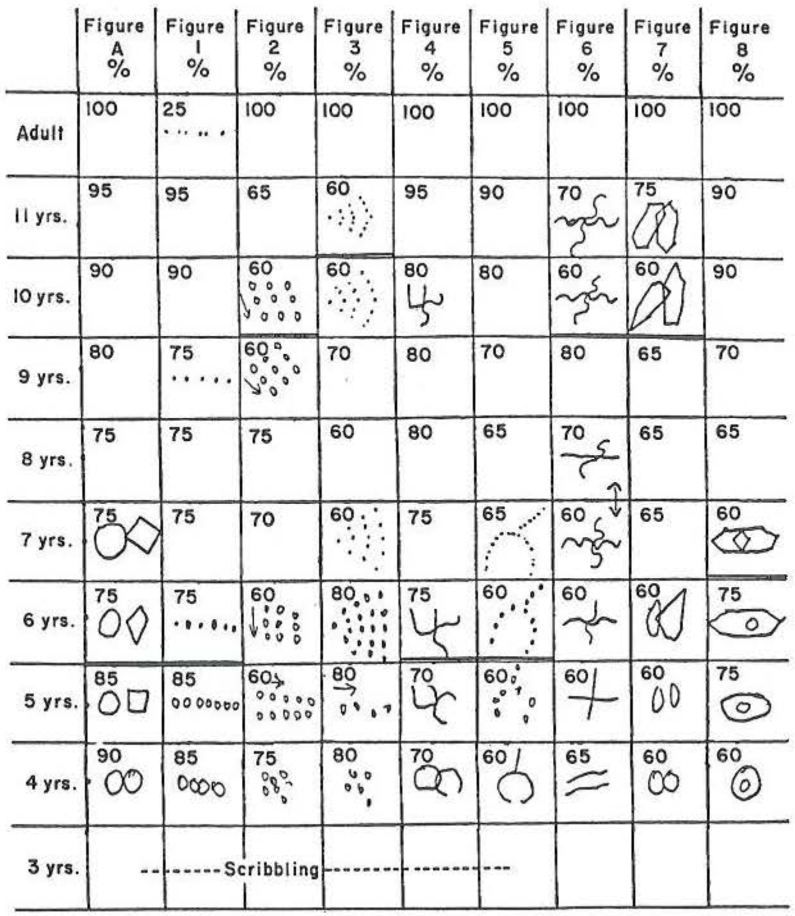


Fig. 50. (From Bender, 1938.)
Summary Chart of Bender's Results with Children.—Percentages in the upper left-hand corner refer to children who could do the type of response depicted or better.

Any such stress on accuracy pure and simple throws away valuable information relating to the basic maxim projective testing was found to share with Gestalt psychology, that is, that deviation from one thing also may be assimilation *toward* something else. Indeed, there is a whole array of saliently divergent possibilities of "improvement," each of which involves increasing the error in a different manner, and thus at the same time defines deterioration. Each of these possibilities may be laden with potential diagnostic significance of its own, and much information is likely to be lost in lumping them all together.

The ontotropic bias of the Rorschach stimulus materials should not prevent

the clinician from attempting to establish a truly Gestalt-psychological formalism subscore on this test along with the one suggested above for the Bender; some of the ecological imbalance of the Rorschach could thus be rectified. Existing so-called form scores on the Rorschach are somewhat oblique to this task; they are based predominantly on what in Sander's terminology would appear to be realistically meaningful rather than geometric-ornamental criteria of classification. Contrariwise, a comprehensive ontotropism score could also be developed for the Bender in spite of the fact that its stimulus materials offer a minimum of invitation for structuring in that direction. But it would be significant to know whether or not a person will, despite the existing stimulus handicaps, structure formalistically on the Rorschach or realistically on the Bender. In this manner the representativeness of research could be increased by purely evaluative means without altering the stimulus.

4. THE NUMBER OF UNIVERSES TO BE SAMPLED

Aside from the subjects, on the one hand, and the physical or social objects, on the other, there are several other universes that are being, or will in the end have to be, sampled representatively. William Stephenson (1953) and Carl Rogers (1952) have randomly sampled traits or types of responses, although perhaps from somewhat arbitrarily constituted trait-universes which perpetuate the bias of their original constitution. Most certainly, attitudes should be sampled. Many kinds of response variables could and should be sampled.

Merely as a terminological policy, we should like to suggest that the phrase "representative design" be confined to cases in which representativeness is sought for elements of the external, ecological situation. Repeated applications of a test (intraindividual differences), the sampling of attitudes, statements, adjective traits, or behaviors from a "repertory" universe, all these must be classified with subject sampling rather than with object sampling. Needs toward ecological representativeness are so urgent, past neglect so drastic, and the inhibitions of tradition so powerful that our methodological watchword should be kept as sharply focused as it was originally intended to be.

Representative design, then, is but a special case of representativeness at large as we know it to be a universal research requirement. It is geared to the way in which situations occur in life. Each situation is a "variate package," that is, a more or less incidental combination of specific values along a large, and indeed unknown, number of dimensions. Ecologies, and the situations that constitute them, are in many ways like persons, which also are variate packages. Ecologies or situations exhibit consistencies and "habits" all of their own, although perhaps less strikingly than do individuals; we may "know" them and like or dislike them as we do our fellow men.

It is by virtue of these relative consistencies that variate packages as a whole, and not their isolated dimensions, should be taken to define a universe.

If we accept this position, there are only two major types of universes, the responder-populational and the ecological, as we have assumed them in Part One. As soon as we start ignoring the package character of individuals (including their traits, responses, attitudes, etc.) and of situations, more universes must be distinguished, however. If we hear, as we often do, that a "variable" has been sampled, and do not at the same time include its contingent associates in other dimensions, or disrupt these relationships, there must be as many universes—and as many tests of significance—as there are variables. In factorial and related forms of systematic design this stage of artificial disintegration is taken as the model case.

On the other hand, we may also think of person-attitude-situation complexes as compound packages and sample these. An example of this lumping together of universes would be if each of the 93 situations in our size-constancy survey (§ VII/3) had been occasioned, and responded to, by a different person instead of by the same individual. In this manner the single test of significance envisaged by Jarrett (see footnote 15a, p. 39) as encompassing both the responders and the ecology may yet become possible.

Conclusion: Functional Theory and a Definition of Perception

As we have said, the distal constant functions established by the perceptual system reflect a generalization achievement on the part of the organism. In turn, the wide-arching functional validities as they appear in figures 5 and 10, and—within systematic designs—in all constancy ratios, constitute the corresponding generalized statements by the researcher regarding the organism's perfection in the attainment of a given distal variable. The respective coefficients or ratios thus must be classified as cases of "descriptive theory" by virtue of their generality alone ("psychology in terms of objects," see Brunswik, 1934, 1936, 1937).

Our canvassing inventory or mapping of the array of outpost variables attained—including the degree of such focusing—has led us to further generalizations concerning preferential focusing on distal rather than proximal variables at large. This opens the way toward a general theory concerning the relationship between life-relevance and distality and concerning problems of biological-functional economy.

As is indicated especially in figures 6, 7, and 8, each over-all functional arc or achievement may be broken down into an extrasystemic and an intrasystemic constituent; these constituents have been called ecological validity and utilization, respectively. The general pattern of the mediational strategy of the perceptual system is predicated upon the limited ecological validity or trustworthiness of cues which we have observed in many contexts throughout this book. The limitations in the dependability of single-cue variables force an uncertainty-geared probabilistic strategy upon perception. In order to improve the cognitive "wager"—as Reichenbach (1938) would label this (see p. 56)—the perceptual system must accumulate and combine cues. Thus we arrive at a more complete understanding of the principle of mutual substitutability or "vicarious functioning" of means (or of cues) which Hunter,

Tolman (1932, chap. i) and most other behaviorists looking for a structural criterion have incorporated in their basic definitions of behavior or purpose (see Brunswik, 1952, chap. ii). The lens-like model in figure 10 summarizes the relationship between achievement and strategy and may be taken to exemplify the basic unit of psychological functioning. Regardless of how much the attainment is improved, however, distal function remains inherently probabilistic.

In the light of this functional model all "constant"—or rather quasi-constant—function, be it "intuitive" or explicit, can be explicated as a statistical reasoning process remindful of Helmholtz's "unconscious inference." An important difference arises from the fact that, as we have seen, the introspectionistic and perfectionistic overtones of the Helmholtzian doctrine are no longer tenable; neither can it be assumed that the premises and the derivation must originally have been given in consciousness, nor is the type of intuitive reasoning found in perception of a penetrating or foolproof nature. Forming a Latin-Greek hybrid, we therefore prefer to speak of our view as a mere "ratiomorphic" rather than a fully rational model of achievement (Brunswik, 1954, 1955). Not only is ratiomorphism not to be confused with rationalism or with intellectualism; as we have seen especially in § XIV, it even helps us to nail down more concretely the rather important secondary differences between "perception" and "thinking."

One of the most important aspects of functional theory concerns the relationship between ecological validity and utilization. Ideally, cues should be utilized in accordance with their validity; as we have seen, the organism tends to do so to a certain extent, even though there may be important exceptions (for the latter see Hammond, 1955, reporting on the work of his collaborators at Colorado; see also a study by Smedslund (1955) at Oslo). Here we must further inject the element of "cost" to the organism—e.g., in developing the integrated "double-eye" for the binocular cues of depth—just as we must ask for the cost of an automobile together with its expected return in budgeting our expenditures. In this respect functional theory takes on certain features of economic theory.

While the utilization of cues has long been a favorite topic of psychology, the quantitative study of ecological validity in the framework of a textural ecology is relatively new. Outside of psychology, a first step in the direction of the bivariate type of correlation analysis which is covered by our concept of ecological validity has been made by Wiener (1949) in his analyses of double time series in the general framework of cybernetics. These analyses furnish a second point of affinity between functional theory in psychology, on the one hand, and economic theory, on the other. Double time series are, as Wiener points out, most conspicuous in economico-sociological and meteorologico-geophysical applications; in both instances the relative lead of one time series with respect to another may well give much more information concerning the past of the second than of its own. For example, on account of the general eastward movement of the weather, Chicago weather may be more important in the forecasting of Boston weather than Boston weather itself. It will be noted that even here the comparison stays within the same kind of variable or physical denomination ("weather"); in all cases of this kind the comparisons or correlations involved have not yet reached the stage of genuinely bivariate analysis such as is given by correlating psychological cues with object variables, or means with goal-attainments.

A more direct exposition of those of the mathematical principles of communication which are of particular relevance to the understanding of focusing by vicarious functioning as it occurs in psychological mechanisms has been given by Shannon and Weaver (1949). In terms of the vocabulary of the special brand of telecommunication engineering involving semi-controlled media to which Shannon's theory has been geared, perceptual cues and behavioral means are like "signals" in "coded messages." The mediating channels are contaminated with interferences or constraints of their own. The result is equivocation. It is then "not in general possible to reconstruct the message with certainty by any operation on the signal." Shannon's diagram showing the fanning-out of "reasonable causes" (messages, inputs) for a given "high-probability received signal" or effect, and of "reasonable effects" (signals, outputs) from a given "high-probability message" or cause in a channel, bears formal resemblance to the equivocal types of coupling between intra- and extraorganismic regions to which this writer has called attention some time ago (1934, fig. 2) and which can also be read into the special application of the lens model presented in our present figure 10.

Whenever the "capacity" of a channel is less than the richness of variability of the source from which it accepts messages, the channel is "overloaded." In this case no code will reduce the error frequency as low as one may please. Shannon and Weaver point out that regardless of how clever one is with the coding process, it will always be true that after the signal is received there remains some undesirable (noise) uncertainty about what the message was.

The crux of organismic adjustment which we have studied in this book may be rephrased in quite a similar way: distal perceptual and behavioral mediation must, in the nature of things, in the general case rely on overloaded channels, and the limited dependability of all achievement mechanisms is a result of this overloading. And we must further note that at least part of the trouble lies with the overloading and noise in the external rather than the internal medium.

According to Shannon and Weaver, the chances of error can be decreased by "redundancy," however. Redundancy may be exemplified by, but is by no means restricted to, verbal repetitiveness. When there is noise there is some advantage in not using a coding process that eliminates all the redundancy, for the remaining redundancy helps combat the uncertainty of transmission.

The reader will recognize that the vicariousness of psychological cues and means which we have come to acknowledge as the backbone of stabilized achievement may be viewed as a special case of receiving or sending messages through redundant, even though not literally repetitive channels. The probability of error, given by the variety of possible causes, or effects, that could result in, or be produced by, the type of event in question can thus be minimized. This is the case, for example, in the gain of the over-all functional validity (.99) over the ecological validity of the major retinal cue (.70) in our representative survey of size constancy in which the organism acts as an intuitive statistican (figs. 8, 9, 10).

A suggestion of extending the theory of communication to multivariate patterns of mediation has recently been made by McGill (1953). It is hoped that this will open up the full scope of the vicarious functioning of perceptual cues to formal treatment.

Even so, there is a long way to go from the rather rudimentary emergence of the concept of cue in cybernetics and in the theory of communication to the more varied and somewhat metaphorical applications that would have to be made to render these considerations really fruitful in psychology. Hitherto most of the efforts to apply these disciplines to psychological problems have been rather literal minded and have considered the organism rather than the ecology as the prime source of noise and uncertainty (for details see Brunswik, 1952, 1955; see also Grant, 1954).

The concept of perceptual theory as a ratiomorphic model of functional achievement and of its strategy—that is, as a model of focusing and of vicarious mediation—as well as the attendant methodological postulate of behavior-research isomorphism (Brunswik, 1952) which underlies representative design, are somewhat at variance with the more traditional notions of psychological theory. According to these notions, the main purpose of theory is "explanation" in the sense of a step-by-step tracing of performance to identifiable processes, cues, or tracks of mediation. The ultimate ideal is that of "reduction" of performance to the laws of one of the more microscopic, more fundamental disciplines, notably physiology. The nomothetic behaviorism of Hull (1943*b*) and of his sympathizers, although at first sight of a grossly descriptive nature, reveals its reductionist aim most clearly in the use of a physiologizing terminology.

One of the major devices of the reductive-nomothetic approach is the diacritical confrontation of constant function with its own cues or other mediating instrumentalities. The artificially distortive systematic experiments which exploit the infirming case of an otherwise positive cue, and which range from the stereoscope and pseudoscope to the Ames demonstrations mentioned, serve the same general research aim. To our theory of distal function and of its grand strategy the reductive approach adds the study of tactics. To the study of achievement and of its macro-mediation, both of which fall within the province of functional-representative design, it adds the study of micro-mediation, which falls in the province of nomothetic-systematic design. While the reductive approach reveals the technological details of the machinery that brings about adjustment, representative design and functional approach have their major place in the appraisal of the interplay and relative contribution or weight of environmental factors in the adjustment to a given ecology (see especially p. 55).

The juxtaposition of systematic and representative design which dominates the first half of this book constitutes theory in still another sense. It belongs in that part of the philosophy of science which is best labeled comparative methodology. One of the major concerns of such a theory about policy—or of psychological "metatheory" as Koch (1951) would call it—is with the basic unity of the sciences. In present psychological discussion it is often forgotten that the basic requirement for scientific exactitude is relatively modest and does not include the reductionist-nomothetic-systematic syndrome which some psychologists have adopted under the spell of a somewhat stereotyped image of physics (see Brunswik, 1955). In contrast with this view, we have taken the position that in the end the unity of science is best served by a working out of the thematic diversity of the sciences within the minimum common platform of an objective approach. This diversity of thema involves both the aims of the different disciplines and the research designs capable of serving these aims. The resultant internal unification of psychology centering about the statistical approach has been made clear in the Summary to Part One.

The inextricable entanglement of the methodology of design, on the one hand, and of theory proper, on the other, is not always being realized. The mediating links between the two are the concepts of generality of behavior and of scientific generalization as we have emphasized them throughout our presentation. One who understood this entanglement is Hammond (1951) when he suggested that the situation created in psychology by representative design may be set in analogy to that in relativity physics.[24] In both cases regularities considered universal are revealed as contingent upon a limited ecology. As Hammond points out, universality of law presupposes homogeneity of the universe. Only under this condition does it matter little where and when and over how wide a range of circumstances a phenomenon is studied. Only under this condition may experimental design safely be left to the convenience and liking of the experimenter and thus become systematic. But ecologies are not in all respects homogeneous within the universe.

The relation between physical law and ecological correlation has been illustrated in figure 10. The derivation of the well-known proportionality law of physical optics (center enclosure in the left half of the figure) is possible not only from the ecological validities shown but from all alternate ecologies (as exemplified by our main sample of 93, vs. the subsample of 59 situations for which values appear in parentheses). It is therefore evident that textural ecology adds valuable probabilistic information to the vastly distilled nomothetic information provided by physics proper. As the laws of physics, ecological correlations are abstracted and summarized from the geographies that make up the respective universes. Since ecological universes are more specific than is *the universe* of physics, they are best capable of furnishing the type of information which finite organisms are able to

[24] Hammond has been accused of confusing design and theory; this writer has defended Hammond's position in a Note (1951) in which their inseparability within functional theory and the doctrine of representative design was pointed out.

absorb in learning, notably in probability discrimination with partial reinforcement.

For the reasons stated, the laws of triangulation as applied to the external situation are of somewhat academic interest so far as the penetrative capacities of the intuitive perceptual system are concerned. In the history of psychology the binocular depth mechanism has frequently been absolutized as a reflection of immutable physical law; only recently has it once more been the starting point for nomothetic treatment (Luneburg, 1947; see also Graham, 1951). As always in functional psychology, the question is not one of the "existence" of general law, but only one of its accessibility to the perceiver with his inherent lack of control over the relevant parameters and conditions of application throughout the external medium. The limited ecological validity of the binocular cues within our cultural ecology is obvious in view of the routine exposure of the perceiver to optical instruments as well as to flat pictures as a substitute access to three-dimensional reality.

Surveying what we have said about perception and casually adding some of what was not specifically discussed in this presentation, the perceptual system appears as a complex instrument aiming at a mapping of the distal environment into the organism (§§ V to VIII, XI to XV). Since perception is "persistent" and "docile" (§ XVII) in pursuing this aim, it fulfills Tolman's criteria for "purposiveness" of behavior (1932, chap. i). Its study falls within the scope of the objective approach and requires no recourse to consciousness even though it remains true that the use of so-called introspective or phenomenological methods may expedite its progress.

As to strategy, perception is too limited in scope to abandon the relatively primitive and superficial ("stupid") uncertainty-geared approach, with its competing multiplicity of stereotyped cues of limited ecological validity, in favor of the certainty-geared approach of measurement and calculation with its tendency to fall back on a single cue for each constituent variable. One of the probable results of this rivalry is confluxion interaction (§ IV).

As to attainment, the consequence is compromise and a falling short of precision (§§ V, VII/3, XI–XIII), but also the relative infrequency of drastic error (§ XV). Under highly safeguarded conditions resembling physical measurement (classical tied-variable design, § II) high precision may be exhibited but this precision is of limited ecological generalizability (§ III).

As to promptness and richness, perceptual actuality reaches considerable organizational maturity within fractions of a second (§ XVIII), Katz's (1911) claim nothwithstanding that perceptual constancy may reach its peak as late as three seconds after the onset of observation. At the same time the richness in detail of perception is staggering, even though McCulloch (1955, p. 37 f.) claims a ceiling of "a hundred bits [of incidental information] per second of sustained reception." The ability of perception to organize information into cognitively parsimonious units and subunits in overviewing the situation is unmatched among psychological functions. Its speed and richness make up for much of the shortcomings in strategy and attainment.

As to attitude, perception is relatively fixed when compared with still higher cognitive functions. Habitually it tends to be set toward the biologically more relevant distal and social (physiognomic) variables (§§ XI to XV), even though shift spans may readily be increased by proper training (§ XIII/2) and are of considerable magnitude at higher levels of the hierarchy of perceptual performances (§ XIII/3).

A certain degree of autonomy of perception relative to thinking is evident in some of the data presented, notably in those on prolonged familiarity as affecting the influence of values on perception (§ XII/3), in those on differences of error distribution and related changes when a critical attitude is taken (§ XV), and in those on the self-sufficiency and lower developmental status of perceptual learning (§ XVII).

Perception, then, emerges as that relatively primitive, partly autonomous, institutionalized, ratiomorphic subsystem of cognition which achieves prompt and richly detailed orientation habitually concerning the vitally relevant, mostly distal aspects of the environment on the basis of mutually vicarious, relatively restricted and stereotyped, insufficient evidence in uncertainty-geared interaction and compromise, seemingly following the highest probability for smallness of error at the expense of the highest frequency of precision.

BIBLIOGRAPHY

ALLPORT, G. W. *Personality*. New York: Holt, 1937.

AMES, A. Binocular vision as affected by relations between uniocular stimulus-patterns in commonplace environments. *Amer. Jour. Psychol.*, 1946, 59: 333–357.

ANDERSON, L. D. Estimating intelligence by means of printed photographs. *Jour. Applied Psychol.*, 1921, 5: 152–155.

ANSBACHER, H. Perception of number as affected by the monetary value of the objects. *Archives of Psychol.*, 1937, No. 215.

ARNHEIM, R. *Art and Visual Perception*. University of California Press, 1954.

ATTNEAVE, F. Some informational aspects of visual perception. *Psychol. Review*, 1954, 61: 183–193.

BAXTER, B. Problems in the planning of psychological experiments. *Amer. Jour. Psychol.*, 1941, 54: 270–280.

BENDER, L. A visual motor Gestalt test and its clinical use. *Res. Monogr. No. 3, Amer. Orthopsychiat. Assn.*, 1938.

BENUSSI, V. Zur Psychologie des Gestalterfassens. In: Meinong, A., *Untersuchungen zur Gegenstandstheorie und Psychologie*. Leipzig: Barth, 1904.

BEYRL, F. Über die Grössenauffassung bei Kindern. *Zeitschrift f. Psychol.*, 1926, 100: 344–371.

BILLINGSLEA, F. Y. The Bender-Gestalt: An objective scoring method and validating data. *Jour. Clin. Psychol.*, 1948, 4: 1–27.

BINET, H. La Mesure des illusions visuelles chez l'enfant. *Rev. Phil.*, 1895, 40: 11–25.

BORING, E. G. *History of Experimental Psychology*. New York and London: Century, 1929.

———. *Sensation and Perception in the History of Experimental Psychology*. New York: Appleton-Century, 1942.

BRUNER, J. S., and GOODMAN, C. C. Value and need as organizing factors in perception. *Jour. Abnormal and Social Psychol.*, 1947, 42: 33–44.

BRUNSWIK, E. Zur Entwicklung der Albedowahrnehmung. *Zeitschrift f. Psychol.*, 1928, 109: 40–115.

———. Über Farben-, Grössen- und Gestaltkonstanz in der Jugend. *Bericht über den XI. Kongress für experim. Psychol.*, ed by H. Volkelt. Jena: Fischer, 1930, 52–56.

———. Die Zugänglichkeit von Gegenständen für die Wahrnehmung und deren quantitative Bestimmung. (No. 1 of: Unters, üb. Wahrnehmungsgegenstände, ed. by E. Brunswik.) *Archiv f. die ges. Psychologie*, 1933, 88: 378–418.

———. *Wahrnehmung und Gegenstandswelt*. Vienna: Deuticke, 1934.

———. *Experimentelle Psychologie in Demonstrationen*. Vienna: Springer, 1935.

———. Psychology as a science of objective relations. *Philos. of Science*, 1937, 4: 227–260. (Errata: 1938, 5: 110.)

———. Das Induktionsprinzip in der Wahrnehmung. *XI° Congrès International de Psychologie*, ed. by H. Pieron and J. Meyerson. Paris: Alcan, 1938, 346 f.

———. The conceptual focus of some psychologiacl systems. *Journal of Unified Science (Erkenntnis)*, 1939, 8: 36–49. (Republished in: *Twentieth Century Psychology*, ed. by P. L. Harriman, Philos. Library, 1946, 49–63.) (Referred to as 1939*a*.)

———. Probability as a determiner of rat behavior. *Jour. Experim. Psychol.*, 1939, 25: 175–197. (Referred to as 1939*b*.)
———. Perceptual characteristics of schematized human figures. (Abstract.) *Psychol. Bulletin*, 1939, 36: 553. (Referred to as 1939*c*.)
———. Thing constancy as measured by correlation coefficients. *Psychol. Review*, 1940, 47: 69–78.
———. Organismic achievement and environmental probability. (Part of: Brunswik, Hull, Lewin, Symposium on Psychology and Scientific Method, University of Chicago, 1941.) *Psychol. Review*, 1943, 50: 255–272.
———. Distal focussing of perception: Size-constancy in a representative sample of situations. *Psychol. Monographs*, 1944, No. 254.
———. Social perception of traits from photographs. (Abstract.) *Psychol. Bulletin*, 1945, 42: 535 f.
———. Four types of experiment. (Abstract.) *Amer. Psychologist*, 1946, 1: 457. (Referred to as 1946*a*.)
———. Points of view. In: *Encyclopedia of Psychology*, ed. by P. L. Harriman, New York: Philos. Library, 1946, 523–537. (Referred to as 1946*b*.)
———. Statistical separation of perception, thinking, and attitudes. (Abstract.) *Amer. Psychologist*, 1948, 3: 342.
———. Note on Hammond's analogy between "relativity and representativeness." *Philos. of Science*, 1951, 18: 212–217.
———. *The Conceptual Framework of Psychology*. (International Encyclopedia of Unified Science, ed. by R. Carnap and C. Morris, Vol. I, No. 10.) University of Chicago Press, 1952.
———. "Ratiomorphic" models of perception and thinking. In: *Proc. 14th Internat. Congress Psychol.*, ed. by N. Mailloux. Montreal: 1954.
———. Representative design and probabilistic theory. *Psychol. Review*, 1955, 62: 193–217; 236–242.
BRUNSWIK, E., and CRUIKSHANK, R. M. Perceptual size-constancy in early infancy. (Abstract.) *Psychol. Bulletin*, 1937, 34: 713 f.
BRUNSWIK, E., and HERMA, H. Probability learning of perceptual cues in the establishment of a weight illusion. *Jour. Experim. Psychol.*, 1951, 41: 281–290.
BRUNSWIK, E., and KAMIYA, J. Ecological cue-validity of 'proximity' and of other Gestalt factors. *Amer. Jour. Psychol.*, 1953, 66: 20–32.
BRUNSWIK, E., und KARDOS, L. Das Duplizitätsprinzip in der Theorie der Farbenwahrnehmung, *Zeitschrift f. Psychol.*, 1929, 111: 307–320.
BRUNSWIK, E., und REITER, L. Eindrucks-Charaktere schematisierter Gesichter. *Zeitschrift f. Psychol.*, 1937, 142: 67–134.
BÜHLER, C. *Kindheit und Jugend*. 3d ed. Leipzig: Hirzel, 1931.
BÜHLER, K. *Die Erscheinungsweisen der Farben*. Handbuch der Psychol., I/1. Jena: Fischer, 1922.
BURZLAFF, W. Methodologische Beiträge zum Problem der Farbenkonstanz. *Zeitschrift f. Psychol.*, 1931, 119: 177–235.
CANNON, W. B. *The Wisdom of the Body*. New York: Norton, 1932.
CATTELL, J. McK. On errors of observation. *Amer. Jour. Psychol.*, 1893, 5: 285–293.
CLEETON, G. U., and KNIGHT, F. B. Validity of character judgments based on external criteria. *Jour. Applied Psychol.*, 1924, 8: 215–231.
COCHRAN, W. G. Testing a linear relation among variances. *Biometrics*, 1951, 7: 17–32.
COHEN, J., and QUINN, B. A nomograph for the Brunswik ratio. *Amer. Jour. Psychol.*, 1946, 59: 146–149.
CROSLAND, H. R., TAYLOR, H. R., and NEWSOM, S. J. Intelligence and susceptibility to the Müller-Lyer illusion. *Jour. Experim. Psychol.*, 1927, 10: 40–51.
CRUIKSHANK, R. M. The development of visual size constancy in early infancy. *Jour. Genet. Psychol.*, 1941, 58: 327–351.
CRUMP, S. L. The present status of variance component analysis. *Biometrics*, 1951, 7: 1–16.

CRUTCHFIELD, R. S. Efficient factorial design and analysis of variance, illustrated in psychological experimentation. *Jour. Psychol.*, 1938, 5: 339–346.
———. The determiners of energy expenditure in string-pulling by the rat. *Journ. Psychol.*, 1939, 7: 163–178.
CRUTCHFIELD, R. S., and TOLMAN, E. C. Multiple-variable design for experiments involving interaction of behavior. *Psychol. Review*, 1940, 47: 38–42.
CURTIS, H. S., and WOLF, E. B. The influence of the sex of the examiner on the production of sex responses on the Rorschach. (Abstract.) *Amer. Psychologist*, 1951, 6: 345f.
DUKES, W. F. Ecological representativeness in studying perceptual size-constancy in childhood. *Amer. Jour. Psychol.*, 1951, 64: 87–93.
EBBINGHAUS, H. *Über das Gedächtnis*. Leipzig: Dunker, 1885.
EHRENSTEIN, W. *Grundlegung einer ganzheitspsychologischen Typenlehre*. Berlin: Junker & Dünnhaupt, 1935.
EISSLER, K. Die Gestaltkonstanz der Sehdinge. (No. 3 of: Untersuchungen über Wahrnehmungsgegenstände, ed. by E. Brunswik.) *Archiv für die ges. Psychol.*, 1933, 88: 487–550.
ESTES, S. G. Judging personality from expressive behavior. *Jour. Abnormal and Social Psychol.*, 1938, 33: 217–236.
FECHNER, G. T. *Elemente der Psychophysik*. 2 vols. Leipzig: Breitkopf und Härtel, 1860.
FEIGL, H. Operationism and scientific method. (In: Symposium on Operationism.) *Psychol. Review*, 1945, 52: 250–259; see especially also the "Rejoinder," 284–288.
FIEANDT, K. v. Dressurversuche an der Farbenwahrnehmung. (No. 7 of: Untersuchungen über Wahrnehmungsgegenstände, ed. by E. Brunswik.) *Archiv f. die ges. Psychol.*, 1936, 96: 467–495.
———. *Uber sehen von Tiefengebilden bei wechselnder Beleuchtungsrichtung*. Helsinki: Psychol. Institute, Univ. of Helsinki, 1938.
FINGER, F. W., and SPELT, D. K. The illustration of the horizontal-vertical illusion. *Jour. Experim. Psychol.*, 1947, 37: 243–250.
FISCHEL, H. Transformationserscheinungen bei Gewichtshebungen. *Zeitschrift f. Psychol.*, 1926, 98: 342–365.
FISHER, R. A., *Statistical Methods for Research Workers*. Edinburgh: Oliver & Boyd, 1925 (8th rev. edition, 1941).
———. *Design of Experiments*. Edinburgh: Oliver & Boyd, 1935 (2d ed., 1937).
FRENKEL-BRUNSWIK, E. Motivation and behavior. *Genet. Psychol. Monogr.*, 1942, 26: 121–265.
———. Intolerance of ambiguity as an emotional and perceptual personality variable. *Jour. Personal.*, 1949, 18: 108–143.
———. Personality theory and perception. In: *Perception: An Approach to Personality*, ed. by R. R. Blake and G. V. Ramsey. New York: Ronald Press, 1951.
FRÖBES, J. *Lehrbuch der experimentellen Psychologie*, Vol. I. Freiburg: Herder, rev. ed., 1923.
———. *Lehrbuch der exp. Psychologie: Neue Forschungsergebnisse als Nachtrag zu Band I*. Freiburg: Herder, 1935.
FULLERTON, G. S., and CATTELL, J. McK. *On the Perception of Small Differences*. Univ. of Pennsylvania Press, 1892.
GARRETT, H. E. *Great Experiments in Psychology*. New York: Appleton, 2d ed., 1941.
GARRETT, H. E., and ZUBIN, J. The analysis of variance in psychological research. *Psychol. Bulletin*, 1943, 40: 233–267.
GASKILL, P. C., FENTON, N., and PORTER, J. P. Judging the intelligence of boys from their photographs. *Jour. Applied Psychol.*, 1927, 11: 394–403.
GIBSON, J. J. (ed.). *Motion Picture Testing and Research*. (Army Air Forces Aviation Psychol. Program Report No. 7.) Washington, D.C.: Government Printing Office, 1947.
———. *The Perception of the Visual World*. Boston: Houghton Mifflin, 1950.
GOBETZ, W. A. A quantification, standardization, and validation of the Bender-Gestalt Test on normal and neurotic adults. *Psychol. Monographs*, 1953, No. 356.

GOTTSCHALDT, K. Über den Einfluss der Erfahrung auf die Wahrnehmung von Figuren. *Psychol. Forschung*, 1926, 8: 261–317, and 1929, 12: 1–87.

GRAHAM, C. H. Visual perception. In: *Handbook of Experimental Psychology*, ed. by S. S. Stevens. New York: Wiley, 1951.

GRANT, D. A. On "The analysis of variance in psychological research." *Psychol. Bulletin*, 1944, 41: 158–166.

———. The discrimination of sequences in stimulus events and the transmission of information. *Amer. Psychologist*, 1954, 9: 62–68.

HALSTEAD, W. C. Biological intelligence. *Jour. of Personality*, 1951, 20: 118–130. (Referred to as 1951*a*).

———. Brain and intelligence. In: *Cerebral Mechanisms in Behavior*, ed. by L. A. Jeffress. New York: Wiley, 1951. (Referred to as 1951*b*).

HAMMOND, K. R. Subject and object sampling—a note. *Psychol. Bulletin*, 1948, 45: 530–533.

———. Relativity and representativeness. *Philos. of Science*, 1951, 18: 208–211.

———. Representative vs. systematic design in clinical psychology. *Psychol. Bulletin*, 1954, 51: 150–159.

———. Probabilistic functioning and the clinical method. *Psychol. Review*, 1955, 62: 255–262.

HEBB, D. O. *The Organization of Behavior*. New York: Wiley, 1949.

HEIDBREDER, E. *Seven Psychologies*. New York: Appleton-Century, 1933.

HEIDER, F. Environmental determinants in psychological theories. *Psychol. Review*, 1939, 46: 383–410.

HELMHOLTZ, H. *Handbuch der physiologischen Optik*, Vol. 3, 1866; transl. from 3d ed. by J. P. C. Southall. Rochester: Opt. Soc. of Amer., 1925.

HERING, E. Der Raumsinn und die Bewegungen des Auges. In: *Handbuch der Physiologie*, ed. by L. Hermann, Vol. 3, Leipzig: Vogel, 1879.

HERTZ, M. Wahrnehmungspsychologische Untersuchungen an Eichelhäher. *Zeitschrift f. vergl. Physiol.*, 1928, 7: 144–194 and 617–656.

HESS, E. H. Development of the chick's responses to light and shade cues of depth. *Jour. comp. physiol. Psychol.*, 1950, 43: 112–122.

———. Binocular vision in the chick. Paper presented at the meeting of the A.A.A.S., Berkeley, December 1954.

HILGARD, E. R. and MARQUIS, D. G. *Conditioning and Learning*. New York: Appleton-Century, 1940.

HOCHBERG, J. and MCALISTER, E. A quantitative approach to figural "goodness." *Jour. Experim. Psychol.*, 1953, 46: 361–364.

HOFSTÄTTER, P. R. Über die Schätzung von Gruppeneigenschaften. *Zeitschrift f. Psychol.*, 1939, 145: 1–44.

———. Ruf und Bestand. *Zeitschrift f. angew. Psychol. und Charakterkunde*, 1940, 60: 64–95.

HOLADAY, B. E. Die Grössenkonstanz der Sehdinge. (No. 2 of: Untersuchungen über Wahrnehmungsgegenstände, ed. by E. Brunswik.) *Archiv f. die ges. Psychol.*, 1933, 88: 419–486.

HOLLINGWORTH, H. L. *Vocational Psychology*. New York and London: Appleton, 1916.

———. *Judging Human Character*. New York: Appleton, 1922.

HOLWAY, A. H. and BORING, E. G. Determinants of apparent visual size with distance variant. *Amer. Jour. Psychol.*, 1941, 54: 21–37.

HOTELLING, H. Analysis of a complex of statistical variables into principal components. *Jour. Educ. Psychol.*, 1933, 24: 417–441 and 498–520. (Also as a separate monograph, Warwick & York, 1933.)

HSIA, Y. Whiteness constancy as a function of difference in illumination. *Archives of Psychol.*, 1943, No. 284.

HULL, C. L. The concept of the habit-family-hierarchy and maze learning. *Psychol. Review*, 1934, 41: 33–54 and 134–154.

———. The problem of intervening variables in molar behavior theory. (Part of: Brunswik, Hull, Lewin, Symposium on Psychology and Scientific Method, University of Chicago.) *Psychol. Review*, 1943, 50: 273–291. (Referred to as 1943*a*.)

———. *Principles of Behavior*. New York: Appleton-Century, 1943. (Referred to as 1943*b*.)

HUMPHREYS, L. G. The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions. *Jour. Experim. Psychol.*, 1939, 25: 141–158.

HUNTER, W. S. *Human Behavior*. University of Chicago Press, 1928.

ITTELSON, W. H. The constancies in perceptual theory. *Psychol. Review*, 1951, 58: 285–294.

IZZET, T. Gewicht und Dichte als Gegenstände der Wahrnehmung. (No. 6 of: Untersuchungen über Wahrnehmungsgegenstände, ed. by E. Brunswik). *Archiv f. die ges. Psychol.*, 1934, 91: 305–318.

JARVIK, M. E. Probability discrimination and the gambler's fallacy in guessing. (Abstract.) *Amer. Psychologist*, 1946, 1: 453 f.

———. Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *Jour. Experim. Psychol.*, 1951, 41: 291–297.

JOHNSON, D. M. A systematic treatment of judgment. *Psychol. Bulletin*, 1945, 42: 193–224.

JONES, H. E. Relationships in physical and mental development. (In: *Mental and Physical Development*.) *Review of Educ. Research*, 1936, 6 (1): 102–123.

JOYNSON, R. B. The problem of size and distance. *Quart. Jour. Experim. Psychol.*, 1949, 1: 119–135.

KARDOS, L. Ding und Schatten. *Zeitschrift f. Psychol., Ergänzungsband 23*, 1934.

KATONA, G. *Organizing and Memorizing*. Columbia Univ. Press, 1940.

KATZ, DAVID. Die Erscheinungsweisen der Farben. *Zeitschrift f. Psychol., Ergänzungsband 7*, 1911. (Quotations refer to abbreviated translation of 2d edition, 1930: *The World of Colour*, London: Kegan Paul, 1935.)

———. Der Aufbau der Tastwelt. *Zeitschrift f. Psychol., Ergänzungsband 11*, 1925.

KENDALL, M. G. *The Advanced Theory of Statistics*. Vol. II. (3rd Ed.) New York: Hafner, 1951.

KINGET, G. M. *The Drawing-Completion Test*. New York: Grune and Stratton, 1952.

KLIMPFINGER, S. Ueber den Einfluss von intentionaler Einstellung und Uebung auf die Gestaltkonstanz. (No. 4 of: Untersuchungen über Wahrnehmungsgegenstände, ed. by E. Brunswik.) *Archiv f. die ges. Psychol.*, 1933, 88: 551–598. (Referred to as 1933*a*.)

———. Die Entwicklung der Gestaltkonstanz vom Kind zum Erwachsenen. (No. 5 of: Untersuchungen über Wahrnehmungsgegenstände, ed. by E. Brunswik.) *Archiv f. die ges. Psychol.*, 1933, 88: 599–628. (Referred to as 1933*b*).

KOCH, S. Theoretical psychology, 1950: An overview. *Psychol. Review*, 1951, 58: 295–301.

KOFFKA, K. *Principles of Gestalt Psychology*. New York: Harcourt Brace, 1935.

KOHLER, I. Über Aufbau und Wandlungen der Wahrnehmungswelt. Vienna: Rohrer, 1951.

KÖHLER, W. *Gestalt Psychology*. New York: Liveright, 1947.

KRECHEVSKY, I. Brain mechanisms and umweg behavior. *Jour. comp. Psychol.*, 1938, 25: 147–173.

KREMENAK, M. Der Eindruckswert der Augengegend auf Grund schematischer Darstellungen. University of Vienna Doctoral Dissertation, 1950.

LAIRD, D. A. *Psychology of Selecting Men*. New York: McGraw-Hill, 1925.

LERSCH, P. *Gesicht und Seele*. Munich: Reinhardt, 1932.

LEVIN, M. M. Weight illusions induced by cues of low validity. (Abstract.) *Psychol. Bulletin*, 1943, 40: 582.

———. Inconsistent position cues in the establishment of perceptual illusions. Univ. of California Doctoral Dissertation, 1946.

LEWIN, K. *Dynamic Theory of Personality*. New York: McGraw-Hill, 1935.

———. Defining the 'field at a given time.' (Part of: Brunswik, Hull, Lewin, Symposium in Psychology and Scientific Method, University of Chicago.) *Psychol. Review*, 1943, 50: 292–310.

Lindquist, E. F. *Statistical Analysis in Educational Research.* Boston: Houghton Mifflin, 1940.
Luckiesh, M. *Visual Illusions.* New York: Van Nostrand, 1922.
Luneburg, R. K. *Mathematical Analysis of Binocular Vision.* Hanover: The Dartmouth Eye Institute, 1947.
McCulloch, W. S. Mysterium iniquitatis of sinful man aspiring into the place of God. *Sci. Monthly,* 1955, 80: 35–39.
McGill, W. J. Multivariate transmission of information and its relation to analysis of variance. *U. S. Human Factors Operations Res. Labs. Report No. 32,* 1953.
Metzger, W. *Gesetze des Sehens.* Frankfurt a.M.: Kramer, 1936. (Second enlarged ed. 1953.)
Mises, R. v. *Probability, Statistics, and Truth.* New York: Macmillan, 1939.
Mohrmann, K. Lautheitskonstanz im Entfernungswechsel. *Zeitschrift f. Psychol.,* 1939, 145: 146–199.
Müller, M. Neue Versuche über die Müller-Lyer'sche Täuschung. Universitiy of Vienna Doctoral Dissertation, 1935.
Müller-Lyer, F. C. Über optische Urteilstäuschungen. *Archiv f. Physiol., Suppl.* 1889, 263–270.
Murray, H. A., and MacKinnon, D. W. Assessment of OSS personnel. *Jour. Consulting Psychol.,* 1946, 10: 76–80.
Pascal, G. R. and Suttell, B. J. *The Bender-Gestalt Test.* New York: Grune and Stratton, 1951.
Paterson, D. G. *Physique and Intellect.* New York: Appleton-Century, 1930.
Pearson, K. *Life, Letters and Labours of F. Galton.* 3 vols. London: Cambridge Univ. Press, 1924.
Pintner, R. Intelligence as estimated from photographs. *Psychol. Review,* 1918, 25: 286–296.
Pintner, R., and Anderson, M. The Müller-Lyer illusion with children and adults. *Jour. Experim. Psychol.,* 1916, 1: 200–210.
Razran, G. H. S. Conditioning away social bias by the luncheon technique. (Abstract.) *Psychol. Bulletin,* 1938, 35: 693.
Reichenbach, H. *Experience and Prediction.* Univ. of Chicago Press, 1938.
Postman, L. The experimental analysis of motivational factors in perception. In: *Current Theory and Research in Motivation.* Lincoln: Univ. of Nebraska Press, 1953.
Richardson, M. W. Multidimensional psychophysics. (Abstract.) *Psychol. Bulletin,* 1938, 35: 659 f.
Rogers, C. R. Case of Mrs. Oak—a research analysis. In: *Studies in Client-centered Psychotherapy,* Vol. III. Washington, D. C.: Psychological Service Center Press, 1952.
Rohracher, H. *Kleine Charakterkunde.* Vienna: Urban and Schwarzenberg, 1952.
Rorschach, H. *Psychodiagnostics.* Berne: Huber, 1942.
Samuels, M. Judgments of faces. *Character and Personality,* 1939, 8: 18–27.
Sander, F. Experimentelle Ergebnisse der Gestaltpsychologie. *Bericht über den X. Kongress für experim. Psychol.,* ed. by E. Becher. Jena: Fischer, 1928, 23–88.
———. Structure, totality of experience, and Gestalt. In: *Psychologies of 1930,* ed. by C. Murchison. Worcester: Clark University Press, 1930.
Schriever, W. Experimentelle Studien über das stereoskopische Sehen. *Zeitschrift f. Psychol.,* 1925, 96: 113–170.
Seidner, S. E. Ecological validity of visual depth criteria. University of California Doctoral Dissertation. (In progress.)
Seiller-Tarbuk, L. Die Eindruckswirkung der Gesichts- und Hauptbehaarung. University of Vienna Doctoral Dissertation, 1951.
Senden, M. v. *Raum- und Gestaltauffassung bei operierten Blindgeborenen vor und nach der Operation.* Leipzig: Barth, 1932.
Shannon, C. E. and Weaver, W. *Mathematical Theory of Communication.* University of Illinois Press, 1949.

Sheehan, M. R. A study of individual consistency in phenomenal constancy. *Archives of Psychol.,* 1938, No. 222.
Shen, E. The place of individual differences in experimentation. In: *Studies in Personality,* ed. by Qu. McNemar and M. A. Merrill, New York: McGraw-Hill, 1942.
Smedslund, J. *Multiple-Probability Learning.* Oslo: Akademisk Forlag, 1955.
Stephenson, W. Correlating persons instead of tests. *Character and Personality,* 1935, 4: 17–24.
———. The foundations of psychometry: four factor systems. *Psychometrika,* 1936, 1: 195–209.
———. *The Study of Behavior.* University of Chicago Press, 1953.
Stern, W. *General Psychology from the Personalistic Standpoint.* New York: MacMillan, 1938.
Strenkovsky, S. *The Art of Make-up.* New York: Dutton, 1937.
Strong, E. K. *Vocational Interests of Men and Women.* Stanford Univ. Press, rev. ed., 1943.
Thomson, G. H. *The Factorial Analysis of Human Ability.* Boston: Houghton Mifflin, 1939.
Thorndike, E. L. Fundamental theorems in judging men. *Jour. Applied Psychol.,* 1918, 2: 67–76.
———. A constant error in psychological ratings. *Jour. Applied Psychol.,* 1920, 4: 25–29.
Thouless, R. H. Phenomenal regression to the "real" object. *Brit. Jour. Psychol.,* 1931, 21: 339–359, and 22: 1–30.
———. Individual differences in phenomenal regression. *Brit. Jour. Psychol.,* 1932, 22: 216–241.
———. Test unreliability and function fluctuation. *Brit. Jour. Psychol.,* 1936, 26: 325–343.
———. Factor analysis in problems of perception. *XI^e Congrès International de Psychologie,* ed. by H. Piéron and J. Meyerson. Paris: Alcan, 1938 (5^e Commission).
Thurstone, L. L. Current issues in factor analysis. *Psychol. Bulletin,* 1940, 37: 189–236.
———. *A Factorial Study of Perception.* Univ. of Chicago Press, 1944.
Titchener, E. B. *Experimental Psychology: A Manual of Laboratory Practice* (4 parts). New York: Macmillan, 1901–1905.
Tolman, E. C. *Purposive Behavior in Animals and Men.* New York: Century, 1932.
Tolman, E. C., and Brunswik, E. The organism and the causal texture of the environment. *Psychol. Review,* 1935, 42: 43–77.
Tryon, R. C. Genetic differences in maze learning ability in rats. *39th Annual Yearbook of the National Society for the Study of Education,* Part I, 1940, 111–119.
Uexküll, J. B. and Kriszat, G. *Streifzüge durch die Umwelten von Tieren und Menschen.* Berlin: Springer, 1934.
Viteles, M. S., and Smith, K. R. The prediction of vocational aptitude and success from photographs. *Jour. Experim. Psychol.,* 1932, 15: 615–629.
Walker, H. M. Degrees of freedom. *Jour. Educ. Psychol.,* 1940, 31: 253–269.
———. *Elementary Statistical Methods.* New York: Holt, 1943.
Wallace, R. P. Apparent personality traits from photographs varied in bodily proportions. (Abstract.) *Psychol. Bulletin,* 1941, 38: 744 f.
———. Some apparent personality characteristics as judged from photographs varied in bodily proportions. University of California Master's Thesis, 1947.
Wartegg, E. Gestaltung und Charakter. *Zeitschrift f. angew. Psychol., Beiheft 84,* 1939.
Weber, C. O. The relation of personality trends to degrees of visual constancy correction for size and form. *Jour. Applied Psychol.,* 1939, 23: 703–708.
Werner, H. *Comparative Psychology of Mental Development.* New York: Harpers, 1940.
Wertheimer, M. Untersuchungen zur Lehre von der gestalt, II. *Psychol. Forschung,* 1923, 4: 301–350.
Wiener, N. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series.* New York: Wiley, 1949.

WINKLER, M. Der Ausdruckswert der Mundgegend auf Grund schematischer Darstellungen. Universitiy of Vienna Doctoral Dissertation, 1951.

WOODWORTH, R. S. *Experimental Psychology*. New York: Holt, 1938.

WOODWORTH, R. S. and SCHLOSBERG, H. *Experimental Psychology*. Revised Ed. New York: Holt, 1954.

WULF, F. Über die Veränderung von Vorstellungen (Gedächtnis und Gestalt). *Psychol. Forschung*, 1922, 1: 333–373. (Translated in: Ellis, W. D. *A Source Book of Gestalt Psychology*. New York: Harcourt Brace, 1939, 136–148.)

WUNDT, W. Über Ausfrageexperimente und über die Methoden zur Psychologie des Denkens. *Psychol. Studien*, 1907, 3: 301–360.

WÜRSTEN, H. *L'Évolution des comparaisons de longeurs de l'enfant à l'adulte*. Neuchatel: Delachaux, 1946.