SYSTEMATIC AND REPRESENTATIVE DESIGN OF PSYCHOLOGICAL EXPERIMENTS

With Results in Physical and Social Perception

EGON BRUNSWIK

UNIVERSITY OF CALIFORNIA, BERKELEY

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*, 1948). 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 anlysis 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 (1948). 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 physi*cist'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 botanics 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 "covariation" among variables in an experiment there are again the two major possibilities of systematic and of representative design.

a) *Systematic covariation.*—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

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.

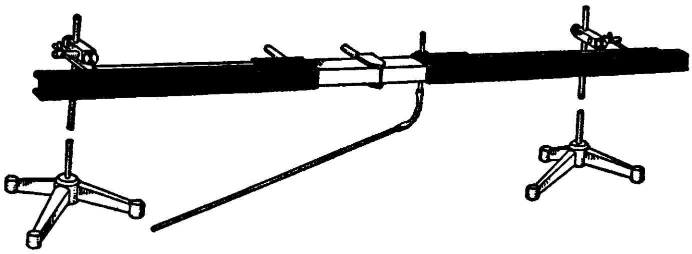


Fig. 1. (Among other sources see Garrett, 1941, fig. 47, A.)

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

"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.

[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

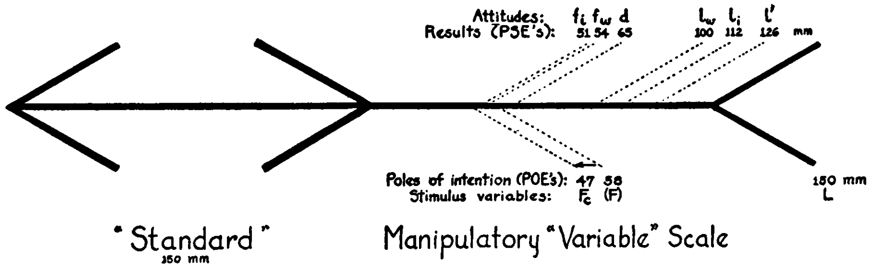


Fig. 2. (Data by M. Müller, 1935.)

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).

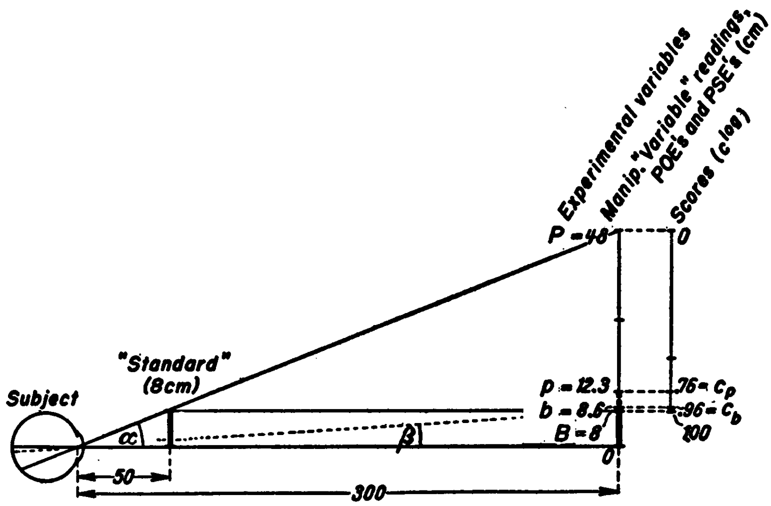


Fig. 3. (Adapted from Brunswik, 1935, fig. 111.)

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.

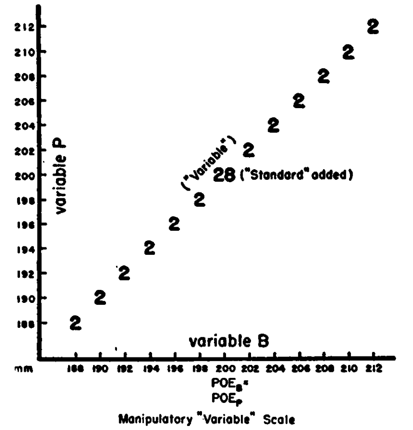
V. Distal Multidimensional Psychophysics of the "Thing-Constancies": Stabilization of Relations with the Remote Environment

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).

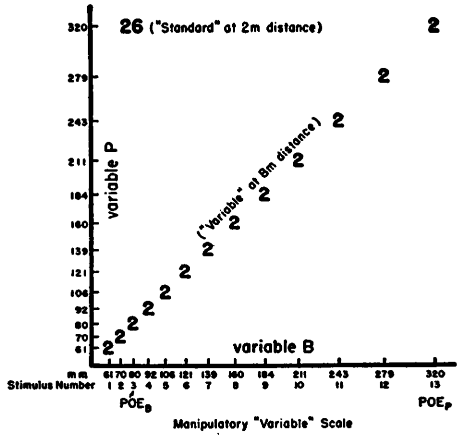
In the size-constancy experiment chosen to represent this step (Experiment C, fig. 3)[10] retinal extensions are pitted against the sizes of objects in the

[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 "overconstancy," 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; 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, 1948.)

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^{los}	
	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-

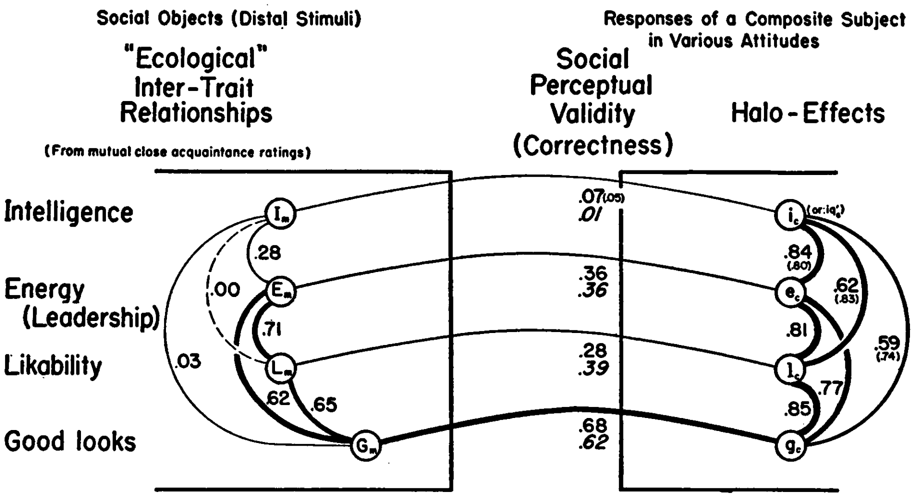


Fig. 5. (Data by Brunswik; preliminary report, 1945.)

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

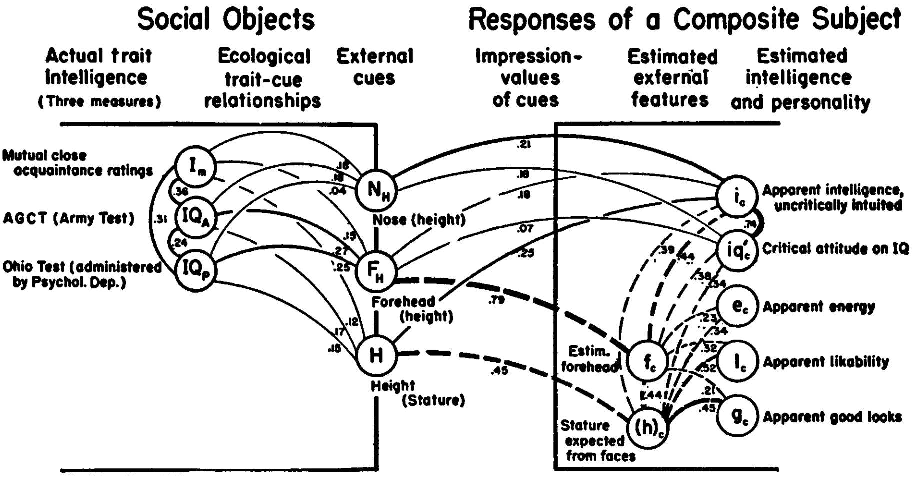


Fig. 6. (Data by Brunswik; preliminary report, 1945.)

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.

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; but 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)
⋮	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
⋮	European—Academic	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, 1939*a*, further specified 1948). 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			(Central-) Organismic variables	Response variables
	Distal stimuli covert	Distal stimuli overt	Proximal stimuli		
	S_C	S_D	S_P	O	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

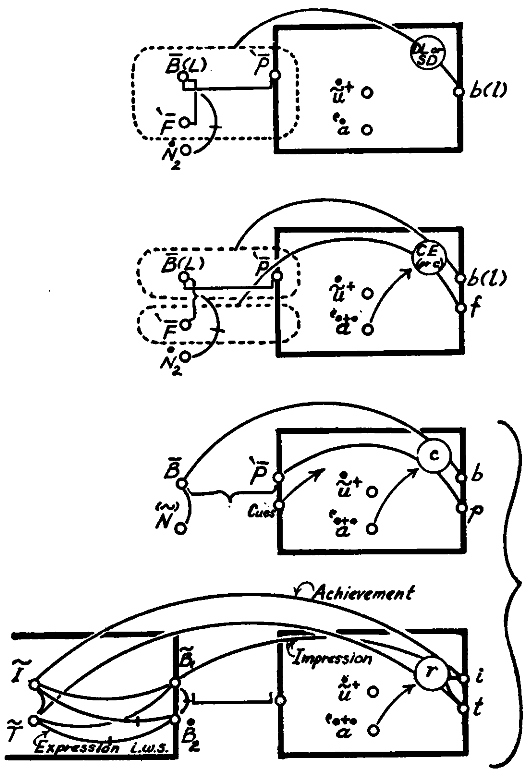
M-crucial alternative: Bodily size (B) vs. visual angle (P)

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'

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)		*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")	($r_{R_iR_j}$)
(4)	Functional problems (in the wider sense)					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)	and					Stimulus variable(s) in situations concerned	Functional validity (General statistical correctness; Achievement)	(r_{SR})
(7)				Measurement or other "independent approach"	Stimulus variable in situations concerned	Measurement repeated at other time(s)	Ecological reliability (of a stimulus)	($r_{SS'}$)
(8)	Psychological ecology					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 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, 1948). 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 acquaintanceship 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 conditions; 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 feeble-mindedness 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 covariation 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 (1939*c*) 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.

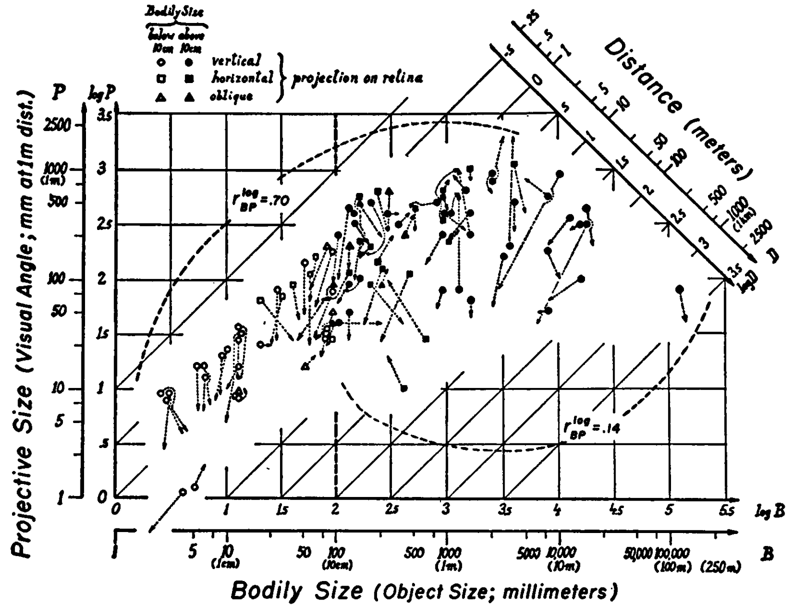


Fig. 8. (From Brunswik, 1944, fig. 2, with coefficients added from table 5.)

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.)

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 "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,

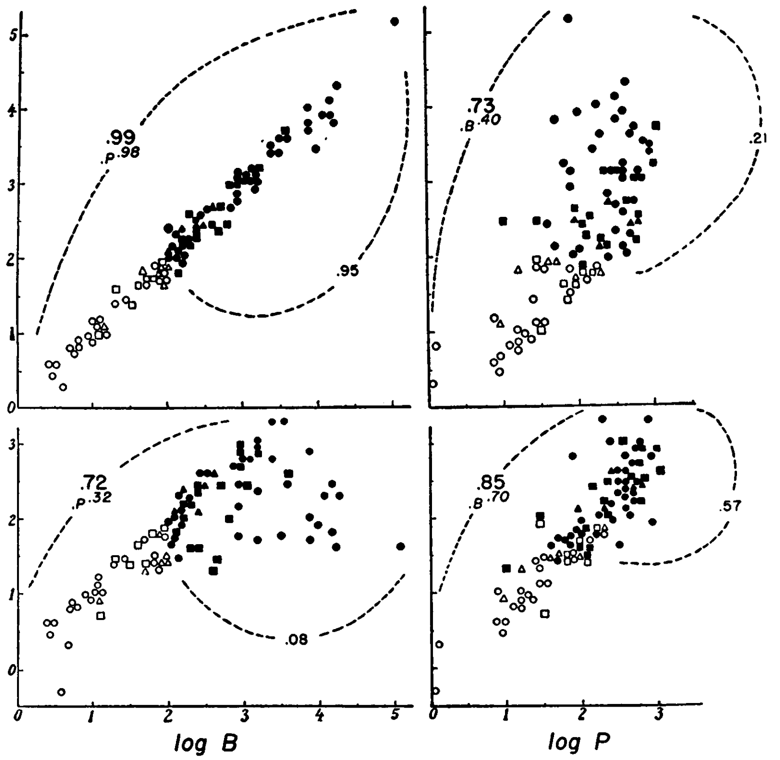


Fig. 9. (From Brunswik, 1944, fig. 3, with coefficients added from table 5.)

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) are again based on logarithms and presented in a manner analogous to figure 8. Partial coefficients are *italicized;* left subscripts preceded by a period mark indicate the variable partialed out.

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).

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, 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 (1948).

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.

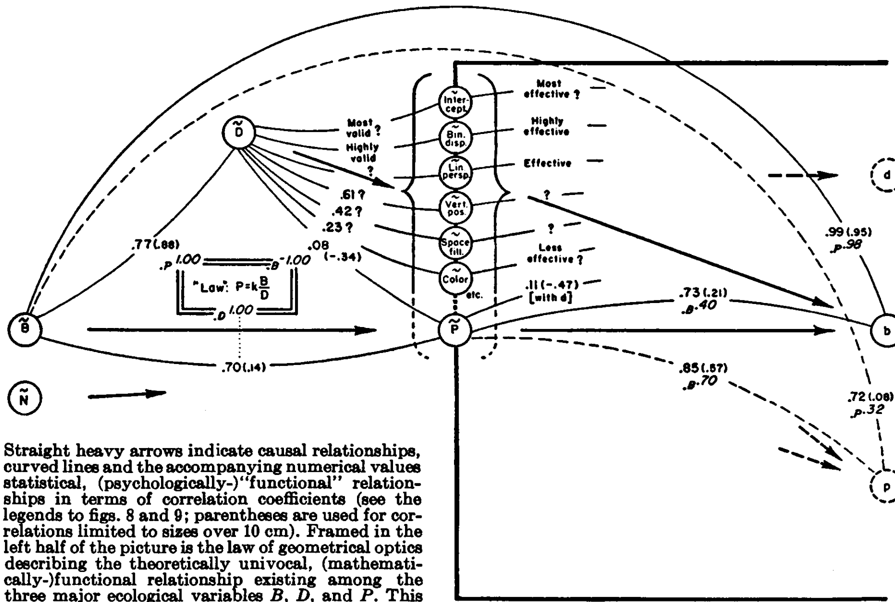
Distal stimulus variables and intra-ecological relationships. Correlation coefficients by Brunswik (1944), rearranged from fig. 8, above.

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.

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 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, 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.

REFERENCES

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

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.

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

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. *XIe Congrès International de Psychologie*, ed. by H. Pieron and J. Meyerson. Paris: Alcan, 1938, 346 f.

———. The conceptual focus of some psychological 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*.)

———. *Methodological Foundations of Psychology.* (International Encyclopedia of Unified Science, ed. by R. Carnap and C. Morris, Vol. I, No. 10.) Univ. of Chicago Press. (In preparation, 1948.)

Brunswik, E., and Cruikshank, R. M. Perceptual size-constancy in early infancy. (Abstract.) *Psychol. Bulletin*, 1937, 34: 713 f.

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, K. *Die Erscheinungsweisen der Farben.* Handbuch der Psychol., I/1. Jena: Fischer, 1922.

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.

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.

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.

Ebbinghaus, H. *Über das Gedächtnis.* Leipzig: Dunker, 1885.

Ehrenstein, W. *Grundlegung einer ganzheitspsychologischen Typenlehre.* Berlin: Junker & Dünnhaupt, 1935.

Fechner, G. T. *Elemente der Psychophysik.* 2 vols. 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.

Finger, F. W., and Spelt, D. K. The illustration of the horizontal-vertical illusion. *Jour. Experim. Psychol.*, 1947, 37: 243–250.

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.

Fröbes, J. *Lehrbuch der experimentellen Psychologie*, Vol. I. Freiburg: Herder, 2d and 3d 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-Century, 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.

Gottschaldt, K. Über den Einfluss der Erfahrung auf die Wahrnehmung von Figuren. *Psychol. Forschung*, 1926, 8: 261–317, and 1929, 12: 1–87.

Grant, D. A. On "The analysis of variance in psychological research." *Psychol. Bulletin*, 1944, 41: 158–166.

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., 1925.

Hering, E. Der Raumsinn und die Bewegungen des Auges. In: *Handbuch der Physiologie*, ed. by L. Hermann, Vol. 3, 1879.

Hilgard, E. R., and Marquis, D. G. *Conditioning and Learning*. New York: Appleton-Century, 1940.

Holaday, B. E. Die Grössenkonstanz der Sehdinge. (No. 2 of: Untersuchungen über Wahrnehmungsgegenstände, ed. by E. Brunswik.) *Archiv für 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.

Hotelling, H. Analysis of a complex of statistical variables into principal components. *Jour. Educ. Psychol.*, September and October, 1933. (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. (In: Brunswik, Hull, Lewin, Symposium on Psychology and Scientific Method, University of Chicago.) *Psychol. Review*, 1943, 50: 273–291.

———. *Principles of Behavior*. New York: Appleton-Century, 1943.

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.

Jarvik, M. E. Probability discrimination and the gambler's fallacy in guessing. (Abstract.) *Amer. Psychologist*, 1946, 1: 453 f.

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.

Katona, G. *Organizing and Memorizing*. Columbia Univ. Press, 1940.

Katz, David. Die Erscheinungsweisen der Farben. *Zeitschrift für Psychologie, Ergänzungsband 7*, 1911. (Quotations refer to abbreviated translation of 2d edition, 1930: *The World of Colour*, London: Kegan Paul, 1935.)

Klimpfinger, S. Ueber den Einfluss von intentionaler Einstellung und Uebung auf die Gestaltkonstanz. (No. 4 of: Untersuchungen über Wahrnehmungsgegenstände, ed. by E. Brunswik.) *Archiv für die ges. Psychol.*, 1933, 88: 551–598.

Koffka, K. *Principles of Gestalt Psychology*. New York: Harcourt Brace, 1935.

Laird, D. A. *Psychology of Selecting Men*. New York: McGraw-Hill, 1925.

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 on 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.

Mises, R. v. *Probability, Statistics, and Truth*. New York: Macmillan, 1939.

Müller, M. Neue Versuche über die Müller-Lyer'sche Täuschung. University 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.

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.

RICHARDSON, M. W. Multidimensional psychophysics. (Abstract.) *Psychol. Bulletin*, 1938, 35: 659 f.

SAMUELS, M. Judgments of faces. *Character and Personality*, 1939, 8: 18–27.

SCHRIEVER, W. Experimentelle Studien über das stereoskopische Sehen. *Zeitschrift für Psychol.*, 1925, 96: 113–170.

SEIDNER, S. E. Ecological validity of visual depth criteria. University of California doctoral dissertation. (In progress.)

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.

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.

STRONG, E. K. *Vocational Interests of Men and Women*. Stanford Univ. Press, rev. ed., 1943.

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 ff.

THOULESS, R. H. Phenomenal regression to the "real" object. *Brit. Jour. Psychol.*, 1931, 21: 339–359.

———. 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. Pieron and J. Meyerson. Paris: Alcan, 1938 (5^e Commission).

THOMSON, G. H. *The Factorial Analysis of Human Ability*. Boston: Houghton Mifflin, 1939.

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.) 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.

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.

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.

WOODWORTH, R. S. *Experimental Psychology*. New York: Holt, 1938.

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.