Birnbaum, M. H. (1975). Expectancy and judgment. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. R. Lindman, & D. B. Pisoni (Eds.), *Cognitive Theory, Vol. 1.* Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 107-118.

5
EXPECTANCY AND JUDGMENT

Michael H. Birnbaum
University of Illinois

Many psychologists seeking explanations of behavior have found judgment to be a rewarding area of investigation. Judgment is a pervasive aspect of human behavior: we say the evening was "pleasantly warm," the man was "tall," his intentions were "not very good," the girl was "extremely beautiful," and she loved him "very much." Judgments even characterize the occupations of physician, lawyer, politician, editor, and the activities of voter, motorist, juror, athlete, and consumer.

The ubiquity of judgmental behaviors identifies judgment as an important problem. More important, perhaps, is the fact that judgments are often public events. They are easily obtained and easily quantified by ordinary rating scales. What meaning one can apply to the numbers obtained in this way is, of course, an important empirical issue (Birnbaum, 1974a,b). But judgment is clearly one area of psychology in which the dependent variable and the problem to be explained are closely allied.

It will be helpful to introduce several empirical themes and methodological ideas that are basic to the present discussion. The first is that judgments are *relative* (Helson, 1964; Parducci, 1968, 1974). So-called "absolute" judgments depend upon the *context,* or stimulus array. The man is judged "tall" because his height exceeds that of a majority of the other men in the context. A 6-ft 2-inch man might be judged "tall" at a cocktail party, but "short" in the context of professional basketball players.

Another fundamental theme is that judgments often require *information integration* (Anderson, 1970b, 1971, 1974a,b); that is, the effects of many stimuli combine to form overall impressions. For example, subjects might be asked to form an impression of a person described as "sincere, intelligent, malicious, and obnoxious." Such a person would probably be judged low on a scale of likeableness (Birnbaum, 1974a). It is assumed that in order to form this impression, the meanings or *values* of the words are combined by an *integration* process to create an impression of likeableness, and this impression is transformed to an overt response ("dislike very much") by a *judgment* process. These processes have been investigated by means of an approach known as *functional measurement* (Anderson, 1970b), in which psychological measurements are derived in accord with the model to be tested. Separation of the stimulus valuation, information integration, and response formation processes can be achieved through extensions of functional measurement methods (see e.g., Birnbaum, 1974a; Birnbaum & Veit, 1974a,b).

A final theme is that man is an active processor of his uncertain, probabilistic environment, an *intuitive statistician* (Brunswik, 1956; Brunswik & Herma, 1951), whose reactions can adapt to changing relationships. Brunswik proposed that the subject's utilization of stimulus cues in perception depends on the predictive value of the cue, and that the pattern of intercorrelations among stimulus events is an important determinant of behavior. For example, a stimulus that produces a trapezoidal retinal image may be either a rectangle viewed from an angle or a trapazoid. The organism must make a "best bet," based on the predictive power of cues.

This chapter views man as dealing with his environment partly through the formation of expectancies that can be compared with actual outcomes or events. The formation of expectancies and the process of comparison are considered fundamental properties of perceptual and judgmental systems. Expectancies can be thought of as the predictions of an intuitive statistician, based on subjective correlations between stimulus cues and events. The combination of cues to form expectancies as well as the comparison of events with expectancies can be studied in terms of information integration. Finally, the relations among events define a context, which can be manipulated to produce predictable effects on judgments.

A THEORY OF EXPECTANCY

Through general experience, organisms are presumed to learn to predict events from correlated cues. As R. L. Gregory and other psychologists have noted, it is useful to conceptualize perception as an *hypothesis* that accounts for sensations. According to this analogy, sensation is to perception as data is to theory. Just as scientific hypotheses yield experimental predictions, perceptions yield sensory predictions.

These predictions or "expectancies" can be thought of as being similar to a statistician's best-fit predictions, since the deviations between perceptual experiences and expectancies should be minimal. Why should the organism form expectancies? It is not enough of an answer to say, *For the same reason that the statistican does,* because the statistician's behavior also deserves explanation.

Perhaps the beginning of an answer can be seen by considering the normative requirements of an organism that lifts objects of varied size, weight, and substance. Suppose you were trying to design a robot that could lift objects in a coordinated fashion. Once designed, the robot is a model of behavior. Of course, the robot should contain a feedback system that allows continuous correction and adjustment of its operations. However, one can also see the advantages of a system that would allow the robot to predict from cues, such as size, the correct forces to apply to move the objects to specified locations. In addition, the robot should be flexible enough to adapt its predictions to a new size–weight relationship in a changing environment.

A statistician could suggest a simple starting point for the robot. It should plot a scatter diagram of weight as a function of size, plotting a separate point for each object. If the objects were all made of the same substance, the correlation would be unity. But in any reasonable environment, such as the set of table-top objects, size and weight will be positively, but imperfectly, correlated. The problem is to predict weight based on size, minimizing the deviations between expectancies and events. The linear regression equation provides a simple solution,

$$\hat{Z}_Y = r_{XY} Z_X , \tag{1}$$

where $\hat{Z}_Y$ is the standard score of the predicted value ($\hat{Y}$), r_{XY} is the correlation between X and Y, and Z_X is the standard score of X. This formula predicts weight as a linear function of size, minimizing the squared deviations between obtained and predicted weight, $(Y - \hat{Y})^2$.

The statistical robot would apply a force of $\hat{Y}$ to an object of size X, and the robot's feedback mechanism would have to correct the force by the amount $(Y - \hat{Y})$. This statistical robot would soon adapt to an environment, requiring the feedback mechanism to do less work as the sample estimate of r_{XY} improved. But this robot would do badly in a changing environment. Suppose the robot were lifting objects of fixed size, all 100 gm in weight. Then $\hat{Y} = 100$ gm. If the robot were then to experience objects all of 200 gm, the robot's expectancy for weight (the average event) would begin to approach 200 as the 200-gm weights are averaged with 100 gm. But it would never reach 200. Even if the robot placed greater weight on more recent events, its predictions would never become unbiased in the new, 200-gm environment. One way the robot could overcome this difficulty would be to form new expectancies by taking a linear combination of the event (Y), the event-expectancy contrast ($Y - \hat{Y}$), and the previous expectancy ($\hat{Y}$). For example, if the robot lifted a 200-gm weight expecting 100 gm, the contrast would be +100 gm. Event (200) plus contrast (100) would be 300 gm; when averaged with the previous expectancy (100), the new expectancy would be exactly 200 gm. Thus, by forming new expectancies based on contrasts with previous events, it may be possible to adapt to a changing environment.

It would seem almost preposterous to propose that humans perform these statistical calculations in their heads or that they could even usefully introspect how they actually form expectancies. The preceding normative considerations provide a set of analogies, a framework for the discussion of perceptual expectancy, adaptation, and judgment. The statistical robot suggests a testable model of expectancy and judgment.

The model presumes that experiences are contrasted with expectancies. Judgments are given by the equation

$$J = E + a(E - E^*), \tag{2}$$

where J is the subjective value of the event, E is what the event would be apart from the expectancy effect, and E^* is the expectancy of the event based on predicting cues. The constant a reflects the magnitude of the contrast effect. Overt ratings are presumed to be a linear function of the subjective values J. Equation (2) gives a qualitative account of the size–weight illusion (Anderson, 1970a, 1972). Since larger objects would be expected to be heavier, they are judged lighter than smaller objects of the same weight. Equation (2) also explains why blind dates are doomed to failure. In order to get a friend to agree to a blind date, E^* must be made large; but the greater E^*, the less the judgment of satisfaction when the friend actually meets the date.

Birnbaum and Veit (1973) proposed that expectancies depend upon the subjective correlation R_{EP} between the event E and the predictor P:

$$E^* = R_{EP} \cdot P, \tag{3}$$

where E^* is the expectancy for the event. Equation (3) is analogous to Eq. (1), except Eq. (3) involves subjective correlation and subjective stimulus values. The subjective correlation is assumed to depend on prior experience and the actual correlation between events in the experimental situation; it need not be equal to the actual correlation in a given situation.

EXPERIMENTAL ILLUSTRATION

A recent experiment done in collaboration with Ken Hagen illustrates the essential ideas of the present theory. Consider the lever in the upper portion of Fig. 1. On each trial, the subject's task is to press at point A, lifting a weight at one of three positions (distances from the fulcrum, B), and judge the force ("effort") required to do so.

If the *same* weight were moved from position 1 to position 3, the actual force required at point A would increase with increasing distance from the fulcrum. However, in this experiment, *different* weights are placed at different locations on different trials to alter the subjects' expectancies. The lower panel of Fig. 1 illustrates the experimental design; each symbol represents a different type of stimulus presentation. The three solid circular points in each row represent different actual weights that, when individually

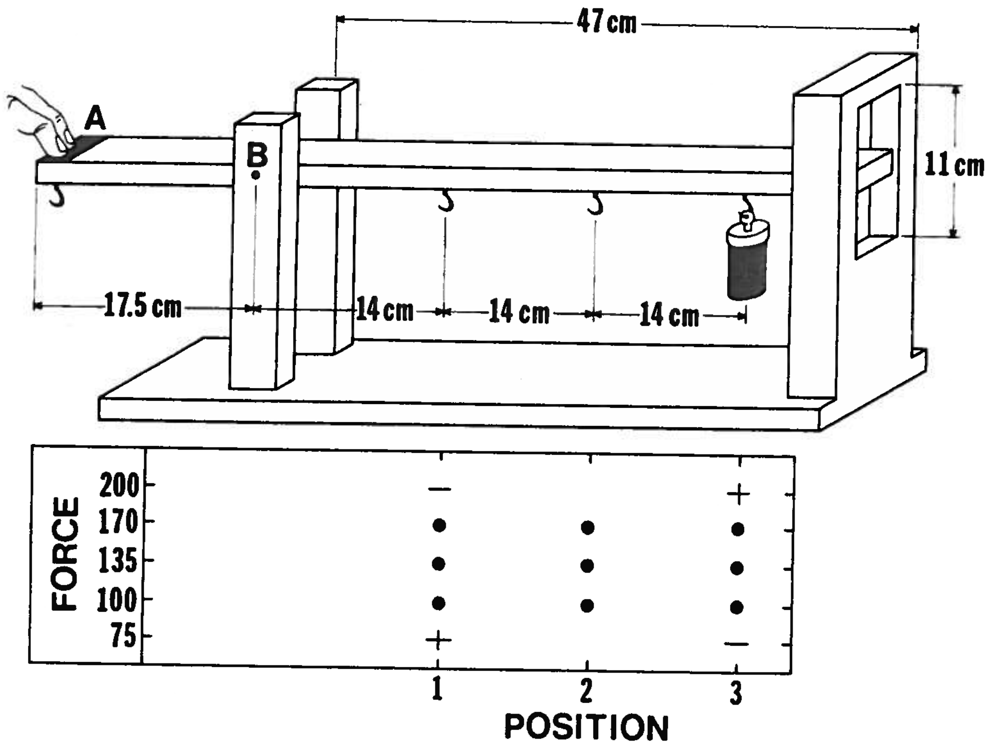

FIG. 1. The lever: The subject presses at point A, lifting a weight at one of three distances from the fulcrum at point B. The task on each trial was to rate effort required to move lever. The lower panel shows the experimental design. Each symbol represents a stimulus presentation. Ordinate values represent the force in gram equivalents that would be required at point A to balance the lever. Solid circles represent test presentations that were the same for all correlation conditions; plus and minus signs represent contextual presentations for the positive and negative conditions, respectively.

placed at the appropriate position, would balance a constant force at point A (indicated by the ordinate value). The weights are all identical in appearance.

Figure 1 also depicts the additional, contextual trials that are presented to change the subjective correlation for different groups of subjects. Subjects in the positive correlation condition are presented with many trials on which a large force (200 gm) is required to lift a weight at the far position or a small force (75 gm) is required for a weight at the near position (see plus signs in Fig. 1). Presumably, after a number of trials they learn to expect greater forces at farther positions. However, subjects in the negative correlation condition are presented with trials that require a large force (200 gm) to lift a weight at the near position or a small force (75 gm) to lift a weight at the far position (minus signs in Fig. 1); consequently, they should expect that objects placed at the nearer positions will require greater forces to lift. Subjects in the zero correlation condition are given both types of presentations. All groups receive the test trials (solid circles in Fig. 1); every other trial is a contextual presentation for the appropriate correlation condition. Thus, the marginal distributions of forces and positions are the same for all groups, but the correlation between force and position is either positive, negative, or zero.

How will judgments of the force required at A depend on position when actual force is held constant? Equation (2) predicts that judgments reflect a contrast between force applied (E) and force expected (E^*), where E is presumed to depend on actual force and E^* is presumed to depend on position. Equation (3) predicts that the direction of the expectancy depends upon the subjective correlation between force and position. In the positive correlation, subjects should expect greater forces at farther positions; hence, judgments should vary inversely with distance. However, subjects in the negative correlation should have the opposite expectancy, and judgments should vary directly with position.

Each panel of Fig. 2 plots mean judgments for a different group (eight subjects), who experienced positive, zero, or negative correlations between position and force. Within each panel, mean judgments of effort are plotted as a function of distance from the fulcrum (position) with a separate curve for each level of actual force (labeled in gram equivalents).

The model predicts that reversing the subjective correlation should reverse the effect of position. Indeed, Fig. 2 shows that judgments vary *inversely* with position in the positive correlation, vary *directly* with position in the negative correlation, and show little effect of position in the zero correlation. When judgments are plotted separately for each subject as a function of position, the data for all eight subjects in the positive condition had negative slopes and for all eight subjects in the negative condition had

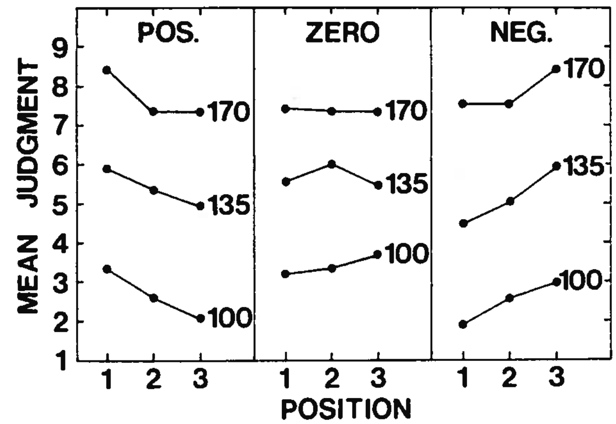

FIG. 2. Mean judgment of effort required at point A plotted as a function of position, with a separate curve for each level of actual force. Each panel shows results for a different group of subjects with a different correlation. When the correlation is positive, judgments vary inversely with position; when the correlation is negative, judgments vary directly.

positive slopes. As predicted by the model, the effect of position can be reversed by manipulating the force–position correlation.

Equation (2) also predicts that the curves in each panel should be parallel, since the differences between the curves are a function of force only. The interaction between force and position, and the three-way interaction with correlation were nonsignificant, in agreement with Eq. (2). These results are consistent with the model, and add further support to the findings of Birnbaum and Veit (1973) and of Birnbaum, Kobernick, and Veit (1974) for the size–numerosity illusion.

Data for the size-numerosity illusion (Birnbaum *et al.*, 1974, Expt. I) are shown in Fig. 3. Judgments of the numerosity of dots vary inversely

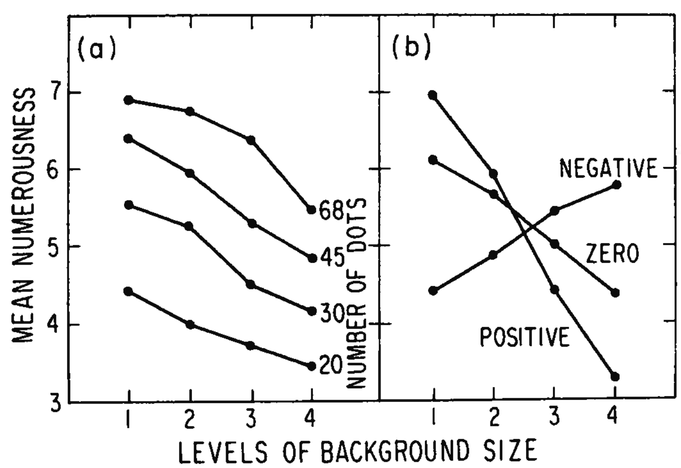

FIG. 3. Results for the size–numerosity illusion: (a) mean judgments of numerosity as a function of the background size with a separate curve for each level of actual number of dots; (b) mean judgments averaged over levels of number, with separate curves for differerent groups who received different size–numerosity correlations. [From Subjective correlation and the size–numerosity illusion, by Michael H. Birnbaum, Marc Kobernick, & Clarice T. Veit, *Journal of Experimental Psychology*, 1974, **102**, 537–539. Copyright by the American Psychological Association. Reproduced by permission.]

with the size of the background on which the dots appear. It is presumed that subjects expect more dots on larger backgrounds. Consistent with Eqs. (2) and (3), when the correlation between size and actual number is negative, the effect of size is reversed.

Birnbaum *et al.* (1974, Expts. I and II) found that subjective correlations estimated from Eq. (3) were monotonically related to the seven actual correlations. When the size–numerosity correlation was actually zero, it was found that the subjective correlation was positive, as if previous experience establishes a positive subjective correlation; that is, as if subjects expect more things to appear on larger backgrounds. In the lever experiment, the subjective correlation (slopes in Fig. 2) is very close to zero when the actual correlation is zero. Expectancies concerning the lever may be relatively weak, or it may be that psychomotor expectancies are labile so that the first series of trials in the zero correlation minimizes the effect of prior experience.

An interesting finding of Birnbaum *et al.* (1974, Expt. II) was that subjective correlations are easily reversed. In relatively few trials, it was possible to reverse the subjective correlation built up at the beginning of the experiment. An interesting problem for further work would be to examine trial-by-trial changes in expectancy and how each type of contextual trial affects subsequent judgments.

SYSTEXTUAL DESIGN

In the present model, the formation of expectancies depends on subjective correlations among events. This empirical proposition is similar to one that led Brunswik (1956) to question the generalizability of systematic research in psychology. It was argued that systematic design obscures the variation and covariation of the independent variables, and if this textured environment is an important determinant of behavior, then the results of systematic research must be ungeneralizable. Brunswik advocated representative design, in which experimental control of the independent variables is given up to achieve sampling of stimulus situations. Each subject is exposed to a context "representative" of the context to which generalization is desired. Poulton (1973) has taken a similar view, but rather than expose each subject to a representative context, he advocated complete between-subject designs to "avoid" context effects.

The present approach is that of *systextual design* (Birnbaum, 1972, 1974b; Birnbaum & Veit, 1973), in which the contextual features of the experimental design are systematically manipulated. Thus, the distributions of the independent variables and their joint distributions (covariances) can be manipulated while maintaining factorial designs (e.g., lower panel of

Fig. 1). This permits investigation of the empirical effects of context while maintaining experimental control.

The lever experiments (as well as the size–numerosity experiments) illustrate a contextual effect of a very general type. The effect of position depends on the experimental design used to measure it. It should be clear that the question, *What is the effect of position?* does not have an answer apart from context. Thus, the important finding is that the direction of the effect depends upon the *pattern* of other stimuli also presented.

Attempts to "avoid" context effects (Poulton, 1973), to sample "representative" contexts (Brunswik, 1956), or even to establish "standardized" contexts are not satisfactory. To investigate the effect of position on judgments of force, a complete between-subject design would use a different group of subjects for each position and force, and could easily find no main effects because judgments are relative. A standardized design might employ a factorial design in which force and position are uncorrelated, and could easily find no effect of position. Representative design would attempt to employ the ecological force–position correlation (positive) and would probably obtain results resembling the positive correlation condition. Between-subject designs simply confound treatments with contexts; standardized or representative designs hold context fixed at some arbitrary value. Hence, these designs do not avoid context effects; nor do they permit assessment of the effects of context. If experimental results are theorized to depend on the context, then systextual design seems appropriate. By manipulating the contextual correlation between force and position, systextual design reveals that the effect of position can be reversed by changing the context.

In another example of systextual design, Birnbaum (1974b) showed that one can use the lawfulness of context effects in psychophysics to derive context-invariant measures. Birnbaum's (1974b) treatment of Parducci's range–frequency theory describes responses as the composition of a context-independent psychophysical function and a context-dependent judgment function. Systematic manipulation of the stimulus distribution led to large context effects that could be separated from the distribution-invariant psychophysical function. It should be emphasized that neither the context-dependent ratings nor context-invariant psychophysical function has greater status, since prediction of judgments requires knowledge of the psychophysical function, the context, and the contextual theory.

Systextual design and functional measurement, together with appropriate theory, can not only lead to context-invariant measures but can also be used to derive measures of the context itself. An interesting feature of the present approach is that one can estimate subjective correlation, a measure of the context that represents an integration of many experiences. Subjective correlation might otherwise seem vague and ephemeral, but it is well-defined in the model and can be estimated by finding the slope of judgments as a function of position and reversing the sign.

ELABORATIONS

The present conception of expectancies is based on the finding that the effect of a cue can be reversed by manipulating its correlation with the judged event. A more general conception of expectancy would take the entire joint distribution into account. Perhaps each stimulus is compared with the distribution of expected events rather than with a single measure of central tendency. The distinction between this more general view and Eq. (2) is analogous to the distinction between Parducci's range–frequency theory (Birnbaum, 1974b; Parducci & Perrett, 1971; Parducci, 1974) and Helson's adaptation-level theory (Helson, 1964; Restle & Greeno, 1970). For example, each event might be judged partly on the basis of its rank in the expected distribution and its relative location in the range of expected values. Equation (2) could thus be tested against the distribution theory by varying the form of the conditional distributions.

An experiment on the size–weight illusion (Birnbaum & Veit, 1974b) has shown systematic deviations from Eq. (2). The effect of weight was less for larger objects than for smaller ones. If objects of all sizes are made of all substances, large objects would be expected to have a greater *range* of weights as well as a greater median weight. Consequently, if objects are judged relative to the expected distribution, the smaller objects should not only be judged heavier, but should also show a greater range of judgments as a function of weight. This interpretation, which could account for the results of Birnbaum and Veit (1974b), could be tested with the lever paradigm by varying the range of forces at different positions. The effect of actual force should be less for positions with greater ranges of weights placed on them. Consistent with this notion, Fig. 2 shows that the effect of actual force is less in the zero correlation condition where the range of weights at positions 1 and 3 is greater. A more convincing test would unconfound range, position, and correlation. This could be accomplished by using four different groups of subjects, who receive only one of the four types of additional contextual presentations.

The lever paradigm may be considered an experimental analogy to perceptual and social judgment. The lever and size–numerosity experiments are directly analogous to the size–weight illusion. More importantly, they are also analogous to a wide array of situations in which expectancies are part of the context that affects judgment. By studying situations in which the relevant variables seem under experimental control, the aim is to uncover basic principles that will have wide applicability for the understanding of judgment. When a theory attains a certain degree of success in a limited domain, its appeal for the discussion and interpretation of other phenomena is enhanced.

ACKNOWLEDGMENTS

This project was initiated at Kansas State University with the support of Bureau of General Research. The author thanks Bonnie G. Birnbaum, Barbara J. Rose, James C. Shanteau, and Clairice T. Veit for comments on an earlier draft.

REFERENCES

Anderson, N. H. Averaging model applied to the size–weight illusion. *Perception & Psychophysics*, 1970, **8**, 1–4. (a)

Anderson, N. H. Functional measurement and psychophysical judgment. *Psychological Review*, 1970, **77**, 153–170. (b)

Anderson, N. H. Information integration and attitude change. *Psychological Review*, 1971, **78**, 171–206.

Anderson, N. H. Cross-task validation of functional measurement. *Perception & Psychophysics*, 1972, **12**, 389–395.

Anderson, N. H. Cognitive algebra: Integration theory applied to social attribution. In L. Berkowitz (Ed.), *Advances in experimental social psychology*, Vol. 7. New York: Academic Press, 1974. (a)

Anderson, N. H. Information integration theory: A brief survey. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology*, Vol. 2. New York: Academic Press, 1974. (b)

Birnbaum, M. H. Systextual design for stimulus variation and covariation. Paper presented at Conference on Human Judgment and Social Interaction, Boulder, Colorado, March, 1972.

Birnbaum, M. H. The nonadditivity of personality impressions. *Journal of Experimental Psychology Monograph*, 1974, **102**, 543–561. (a)

Birnbaum, M. H. Using contextual effects to derive psychophysical scales. *Perception & Psychophysics*, 1974, **15**, 89–96. (b)

Birnbaum, M. H., Kobernick, M., & Veit, C. T. Subjective correlation and the size–numerosity illusion. *Journal of Experimental Psychology*, 1974, **102**, 537–539.

Birnbaum, M. H., Parducci, A., & Gifford, R. K. Contextual effects in information integration. *Journal of Experimental Psychology*, 1971, **88**, 149–157.

Birnbaum, M. H., & Veit, C. T. Judgmental illusion produced by contrast with expectancy. *Perception & Psychophysics*, 1973, **13**, 149–152.

Birnbaum, M. H., & Veit, C. T. Scale convergence as a criterion for rescaling: Information integration with difference, ratio, and averaging tasks. *Perception & Psychophysics*, 1974, **15**, 7–15. (a)

Birnbaum, M. H., & Veit, C. T. Scale-free tests of an additive model for the size–weight illusion. *Perception & Psychophysics*, 1974, **16**, 276–282. (b)

Brunswik, E. *Perception and the representative design of experiments*. Berkeley: Univ. of California Press, 1956.

Brunswik, E., & Herma, H. Probability learning of perceptual cues in the establishment of a weight illusion. *Journal of Experimental Psychology*, 1951, **41**, 281–290.

Helson, H. *Adaptation-level theory*. New York: Harper & Row, 1964.

Parducci, A. The relativism of absolute judgment. *Scientific American*, 1968, **219**, 84–90.

Parducci, A. Contextual effects: A range-frequency analysis. In E. C. Carterette and M. P. Friedman (Eds.), *Handbook of perception*, Vol. II. New York: Academic Press, 1974.

Parducci, A., & Perrett, L. F. Category rating scales: Effects of spacing and frequency of stimulus values. *Journal of Experimental Psychology Monograph*, 1971, **89**, 427–452.

Poulton, E. C. Unwanted range effects from using within-subject experimental designs. *Psychological Bulletin*, 1973, **80**, 113–121.

Restle, F., & Greeno, J. G. *Introduction of mathematical psychology*. Reading, Massachusetts: Addison-Wesley, 1970.