

# Should Contextual Effects in Human Judgment Be Avoided?

E. C. Poulton

*Bias in Quantifying Judgments*  
Hillsdale, NJ: Erlbaum, 1989. 304 pp.  
ISBN 0-86377-105-X. \$49.95

Review by

Michael H. Birnbaum

*E. C. Poulton, affiliated with the Applied Psychology Unit of the Medical Research Council (Cambridge, England), is author of Environment and Human Efficiency, Tracking Skill and Manual Control and The Environment at Work. ■ Michael H. Birnbaum is professor of psychology at California State University, Fullerton.*

Psychophysical judgment is probably the oldest area of psychology, and some people think that its issues have long since been settled. Most introductory psychology textbooks discuss Fechner's logarithmic psychophysical law, present the method of magnitude estimation, and conclude that Stevens's power law describes the relationship between physical and subjective magnitudes. These texts leave the impression that psychologists now agree on how to quantify human judgment and that Stevens' work has been uncontested for the past 30 years. Poulton's book challenges this simple picture by reviewing the factors that influence quantitative judgments. Poulton's (1968, 1979) earlier reviews of this literature are extended and updated in this new volume, whose emphasis is on all of the various ways that an experimenter can get almost any judgment for the same stimulus depending on exactly what is done in the experiment.

The book opens with an overview of how Poulton's perspective differs from that of Stevens (1957, 1975). First, Poulton distinguishes between judgments using "familiar units" (e.g., weight in pounds) and judgments that are based on discrimination among psychological values (e.g., loudness). The book expresses the view, without much evidence, that when judgments are made using familiar physical units, they will be unbiased. Second, whereas Stevens advocated ratio judgments, power functions, and magnitude estimations, Poulton disputes these three aspects of Stevens' work. Instead, he sides with Fechner and argues for a logarithmic psychophysical function (in most cases), difference judgments, and category ratings. These three departures from Stevens are well backed by evidence.

In my opinion, the book's outlook on

contextual effects does not depart enough from that of Stevens, who regarded them as a "nuisance." Poulton's approach is that factors that influence human judgment are "biases" and that biases can and should be "avoided." Unfortunately, the theory of "avoiding bias" is not fully addressed in the book, nor is the philosophy behind it properly developed. The term is explained as follows:

Each chapter discusses the effect that a bias has, and the mechanism that is responsible for it. For simplicity, the name "bias" is used for both the effect and the mechanism. Thus a particular bias is said both to specify a kind of biased judgment and to produce the judgment. (p. 4)

This usage of the term *bias* has four unfortunate consequences. First, the term has a moral connotation that suggests that the experimenter did something "wrong." Second, it evokes an implicit theory that if the experimenter had not "biased" the results, the "truth" would have been revealed. Third, it suggests that once the phenomenon has been named, it has been explained. Poor choices of names might prove resistant to refutation. Fourth, it provides yet another meaning for a term that has already been used quite heavily in other ways. After reviewing the contents of the book, I will contrast the approach of *Bias in Quantifying Judgments* with alternative approaches.

## Summary of biases

In each chapter, there is a simple model of bias, empirical evidence illustrating the bias, and a list of recommended procedures for "avoiding" the bias. Chapter 1 outlines situations that are considered to lead to "valid" or "invalid" judgments. A table in this chapter claims that when the subject is judging physical units of magnitude or subjective intervals (par-

titions) or making linear responses, there is "no problem." However, magnitude estimations of ratios of subjective magnitude are regarded as invalid.

Chapter 2 gives an overview of biases in judgments that are to be covered in the later chapters. Chapter 3 reviews biases in psychophysical methods (tasks). Chapter 4 takes up the idea of subjectively equal stimulus spacing and reviews stimulus spacing and stimulus frequency biases. The basic finding is that if the experimenter presents smaller stimuli with greater frequency, a middle stimulus receives a higher judgment than if it is presented in the context of more frequent larger stimuli. Poulton recommends presenting equally spaced stimuli with equal frequency or presenting each stimulus to a different group of subjects.

Chapter 5 deals with centering biases, in which the same range of responses is centered on the midpoint of two or more stimulus ranges or two or more ranges of responses with different centers are centered on the midpoint of the same stimulus range. These ideas are embedded in the range aspect of Parducci's (1968, 1974) range-frequency theory. Good examples are presented to illustrate these phenomena. For example, a number of experiments by different investigators are reviewed to show that the noise level judged annoying depends on the noise levels of the stimuli presented to the subject. If one wants strict noise standards, one should present subjects with quiet noises only. Under those considerations, the level of noise that would be judged to be "annoying" would be at a lower level than would be inferred in the context of louder noises. The book's suggestion (p. 105) is to use only the stimulus range with the midpoint at which the "judgments are shown to be unbiased" or to present only one stimulus to each subject.

Chapter 6 discusses the logarithmic response bias. The book expresses the view that when subjects judge either subjective values using the procedure of magnitude estimation, judge ratios, judge the loudness of tones, or use a category rating scale with too many categories, then the numerical judgments should be transformed by a logarithmic function (as if the judgments are an exponential function of subjective value). This chapter also reviews the evidence on the ratio/difference controversy and endorses Birnbaum's (1982) conclusion that in many cases, subjects judge differences, despite instructions to judge "ratios" or "differences."

Chapter 7 reviews evidence on "contraction biases," which are akin to regression effects. These biases cause a stimulus to be judged as too close to the average judgment or too close to the previous stimulus or for the standard to be remembered as too close to the average stimulus. The recommendation is to counterbalance orders to try to average out these biases or to present only a single stimulus to each person.

Chapter 8 considers range-equalizing biases. Subjects tend to map the stimulus range onto the response range suggested in the instructions or examples. In magnitude estimation, this effect causes the exponents of fitted power functions to change, depending on the stimulus and response ranges that the experimenter chooses to present. Although such variation is seen as a challenge to Stevens' approach to psychophysics, the book recommends leaving the observers free to choose their own range of responses.

Chapter 9 deals with transfer biases in which other experiences in other situations may influence judgments within the experiment. The book recommends avoiding presentation of example stimuli or responses or using subjects with previous experiences that might affect the responses. Chapter 10 summarizes the factors that influence power functions and concludes that "the case against power functions is that they are not the appropriate way of describing subjective sensory magnitudes" (p. 259).

#### Should bias be avoided?

If the book's greatest strength is its summary of contextual factors that affect judgment, its greatest weakness is its failure to defend its thesis (that "bias" should be avoided) against rival ideas that have been presented in the literature. Poulton assumes that any reasonable person would want to avoid bias. Because "bias" is a negative-sounding, value-laden term, it sounds like something that one should avoid. Procedures are suggested throughout the book to supposedly avoid bias, without much consideration of the idea that we should welcome these so-called "biases" as important data for developing psychological theory.

For example, because pure tones yield results that seemed incompatible with one theory, the book suggests that "the apparent logarithmic response bias in judging pure tones is avoided by avoiding judgments of pure tones" (p. 152). Another approach would be to consider the possibility that the data are correct and that the theory might need revision.

To explore how the philosophy of "avoiding bias" might work in another field, consider the fact that the compass needle does not always point to the azimuth projection of the North Star. Because the North Star is known to change its position slightly during the night, we could refer to this phenomenon as the *North Star bias*. To "avoid" this bias, one might recommend not traveling to places on the earth where the compass and North Star fail to align or obtaining only one measure of North per day. Such procedures, clearly, would avoid only the discovery of disagreements and would not actually fix them.

Suppose contextual effects are true psychological phenomena that would be merely hidden or obscured by attempts to "avoid" them. Do we really avoid contextual effects by avoiding procedures that reveal them? Would it really be best to use only subjects who have never been exposed to a stimulus or used numbers, or are we better off to know exactly how previous experience affects judgments and to use this information in interpreting the data?

It can be argued that there is always some context. Therefore, the suggestion to "avoid bias" amounts to choosing a context. If different results can be obtained in different contexts, then one has the problem of knowing which result is correct before one can recommend the right context. However, if the answer of the investigation is already known, why do it? If one needs to know the proper scaling before one can choose the proper spacing, the philosophy of "avoiding bias" seems circular.

Given the theory that the experimenter's examples should not affect the subject's use of numbers, it might seem proper that the experimenter should not "bias" the subject by giving the subject any ideas. However, from the theoretical viewpoint that the examples are crucial to the subject's behavior, the advice to give the subject "freedom" to choose seems like advising a researcher in a drug test to let the subjects choose whether to take the medicine or placebo without telling the experimenter which one they took. Such "freedom" merely relinquishes control of the experiment to some unknown, possibly confounded factors.

#### Between-subjects designs:

##### Avoiding or confounding context?

Because the stimulus spacing can affect the judgment of each stimulus, Poulton suggests that when the experimenter does not already know the subjective

values of the stimuli, "the only way to be sure of avoiding the stimulus spacing bias is to restrict each group of observers to judging a single stimulus" (p. 95).

However, between-subjects designs confound each stimulus in a different context. In between-subjects designs, it is possible for Group A to assign a higher response to Stimulus A than Group B assigns to Stimulus B, yet a single group of subjects would judge Stimulus B higher than Stimulus A in a within-subject design. Such reversals of rank order (contrasting within- vs. between-subjects experiments) led to a research program that concluded that numbers cannot be compared between different groups of subjects without allowing for different judgment functions in each group (Birnbau, 1982; Mellers & Birnbau, 1982, 1983). This conclusion directly contradicts Poulton's recommendations, but unfortunately this controversy is not addressed in the book.

Birnbau (1982) discussed four approaches to the problem that experimental design can determine the outcome of an investigation. Between-subjects designs attempt to avoid context by having "no context," but they end up confounding the context, the stimulus, and the subject. In a standardized design, the conditions are fixed to specified levels, so that all investigators will obtain the same results. In a representative design, the ecology to which generalization is desired is sampled, and the designs (stimulus distributions and covariations) are chosen to resemble the world to which generalization is desired.

The approach of systextual design is to systematically manipulate the context and to develop a theory to account for its effects. This theory may postulate that some stimulus measures remain invariant with respect to context. For example, Birnbau (1974) derived a psychophysical function from contextual effects using range-frequency theory. He tested whether all of the different stimulus-response functions produced by different stimulus spacings led to the same psychophysical function. The data appeared to satisfy the invariance assumption, and the context-free function so derived also was successful in describing other phenomena (Birnbau, 1982).

#### The psychology of contextual effects

To Parducci (1968), contextual effects in judgment provide a model for understanding the psychology of human experience. Because judgments can be maximized in a negatively skewed distri-

bution, Parducci recommended that one should select a life in which the best events happen often. One should avoid the opportunity to experience a rare and wonderful event if one wants to be happy, according to Parducci, because contextual effects are not merely "biases" in the assignment of numbers but real psychological phenomena of experience. A rare, wonderful event will change how daily life is perceived, not just how it is judged, according to Parducci. Although the book reviews Parducci's experiments on the effects of stimulus frequency and stimulus spacing, it does not address these philosophical issues. The reader familiar with the literature is left wondering where Poulton stands on these issues.

Whether contextual effects can be attributed to scale values that are compared or combined or to judgments is a question that seems akin to these philosophical, phenomenological issues (Mellers & Birnbaum, 1982, 1983). Poulton's treatment of stimulus spacing effects does not address whether the "biases" are attributed to subjective values, responses, or both.

For example, although stimulus spacing affects the rank order of differences in judgment, it may or may not influence the rank order of judgments of "differences." Mellers and Birnbaum (1982) found that stimulus spacing affects differences in judgments but not the rank order of judged "differences" between the same stimuli in the same contexts. These results could be explained by the theory that stimulus spacing within a dimension affects only the judgments but not the scale values that underlie judgments of "differences." Such a demonstration of contextual invariance of scale values (but not judgments) seems to be perhaps an aim of this book, but mysteriously, the book does not advocate ordinal analyses designed to separate judgment functions from scale values.

#### Final comments

This book should be useful for researchers in the field of psychophysics, who will be familiar with Poulton's previous work and the related literature. The book will be difficult reading for those not already familiar with psychophysics, a difficulty that may not be overcome by the book's writing style. The book rarely uses the past tense (even when reviewing early theories and experiments), presents few equations, and intermixes theoretical, empirical, and definitional statements without clear delineation. Graphs, figures, and tables, however, are used ef-

fectively to illustrate the models, data, and suggested procedures.

Overall, the book represents a valuable contribution that will be important to people who plan to quantify human judgment and may not be aware of the many ways in which context can affect the results. For example, anyone who thought that it would be easy to establish standards of annoyance of noises will be annoyed by the number of contextual effects documented so well in Poulton's volume. Although the suggestions for dealing with contextual effects and proper procedures for quantifying judgment will remain controversial, the views set forth in this volume deserve careful study and discussion.

#### References

- Birnbaum, M. H. (1974). Using contextual effects to derive psychophysical scales. *Perception & Psychophysics*, 15, 89-96.
- Birnbaum, M. H. (1982). Controversies in psychological measurement. In B. Wegener (Ed.), *Social attitudes and*

*psychophysical measurement* (pp. 401-485). Hillsdale, NJ: Erlbaum.

Mellers, B. A., & Birnbaum, M. H. (1982). Loci of contextual effects in judgment. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 582-601.

Mellers, B. A., & Birnbaum, M. H. (1983). Contextual effects in social judgment. *Journal of Experimental Social Psychology*, 19, 157-171.

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

Parducci, A. (1974). Contextual effects: A range-frequency analysis. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of Perception* (Vol. 2). San Diego, CA: Academic Press.

Poulton, E. C. (1968). The new psychophysics: Six models for magnitude estimation. *Psychological Bulletin*, 69, 1-19.

Poulton, E. C. (1979). Models for judging sensory magnitude. *Psychological Bulletin*, 86, 777-803.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, 64, 153-181.

Stevens, S. S. (1975). *Psychophysics*. New York: Wiley.