Behavioural Processes 123 (2016) 43–53

Contents lists available at ScienceDirect

Behavioural Processes

journal homepage: www.elsevier.com/locate/behavproc

The problem with categorical thinking by psychologists

Michael E. Young

492 Bluemont Hall, Department of Psychological Sciences, Kansas State University, Manhattan, KS 66506-5302, USA

ARTICLE INFO

Article history:
Received 24 June 2015
Received in revised form
18 September 2015
Accepted 21 September 2015
Available online 28 September 2015

Keywords:
Continuous
Categorical
Statistics
Modeling
Learning
Multilevel analysis
Sampling

ABSTRACT

Continua abound in the natural world, but the treatment of these continua in the psychological study of behavior is often categorical. Four practices and their consequences are considered: limited sampling of a continuum, turning continuous predictors into categories through dichotomization and similar practices, treating continuous predictors as categorical in an analysis, and overuse of the analysis of variance. Concrete examples illustrate the consequences of these practices. Recommendations to improve the treatment of continua, both in the design and analysis of data, are provided.

1. Introduction

Stevens (1946) related a story about a task set before the British Association for the Advancement of Science in 1932–the quantification of sensory events. Eight years later, the committee prepared a final report that documented their continuing disagreement on the issue. It was in the aftermath of their disparity of opinions that Stevens developed a taxonomy of measurement scales in terms of nominal, ordinal, interval and ratio scales and noted that a variable's classification would importantly determine the types of statements that could be made about it and the appropriate statistics for its analysis.

Just before the British Association meeting began its 8 year task, Kurt Lewin (1931) published a seminal paper on the need to move from Aristotelian approaches to psychology to Galileian approaches. He drew on the history of physics to note that "In modern quantitative physics dichotomous classifications have been entirely replaced by continuous degradations" (p. 144). He was optimistic that the beginnings of such a move was occurring in psychology with the advent of sensory psychology and the identification of functional relationships that gave rise to the psychophysical laws that we know today (Fechner, 1860/1966; Stevens, 1957).

In the decades since these publications, however, the design of the typical psychology experiment has fallen into a predictable pattern: a set of nominal predictors and a continuous (interval/ratio) outcome. Although some domains of psychological investigation commonly examine many points along a predictor scale (e.g., psychophysics), a great many published papers in the field (including many of my own) handicap the interpretation and analysis of their experiments by treating potentially continuous predictors as categorical. Statements allowed for a nominal scale predictor are much more restricted than those for interval or ratio scale predictors (Stevens, 1946), thus designs that treat continuous predictors categorically fail to take full advantage of the predictive continua at their disposal and prevent the development of the types of functional relationships that Lewin (1931, 1936) hoped would emerge from psychology. Upon reflection, I noticed this categorical bias was present in much of my own work along with a clear shift over the past 10 years toward incorporating more continua. This observation led me to speculate on the consequences and determinants of this shift.

Not surprisingly, design and statistical innovations that help us to deal with continuous predictors are regularly emerging. These novel approaches appear in a range of journals, textbooks, and statistical software and are put into practice in universities and laboratories across the world, although their penetration is often quite uneven. Despite this progress, the use of these improvements in published papers and their appearance in university curricula is often slow (Aiken et al., 2008; Fidler et al., 2004; Sharpe, 2013).

E-mail address: michaelyoung@ksu.edu

http://dx.doi.org/10.1016/j.beproc.2015.09.009
0376-6357/

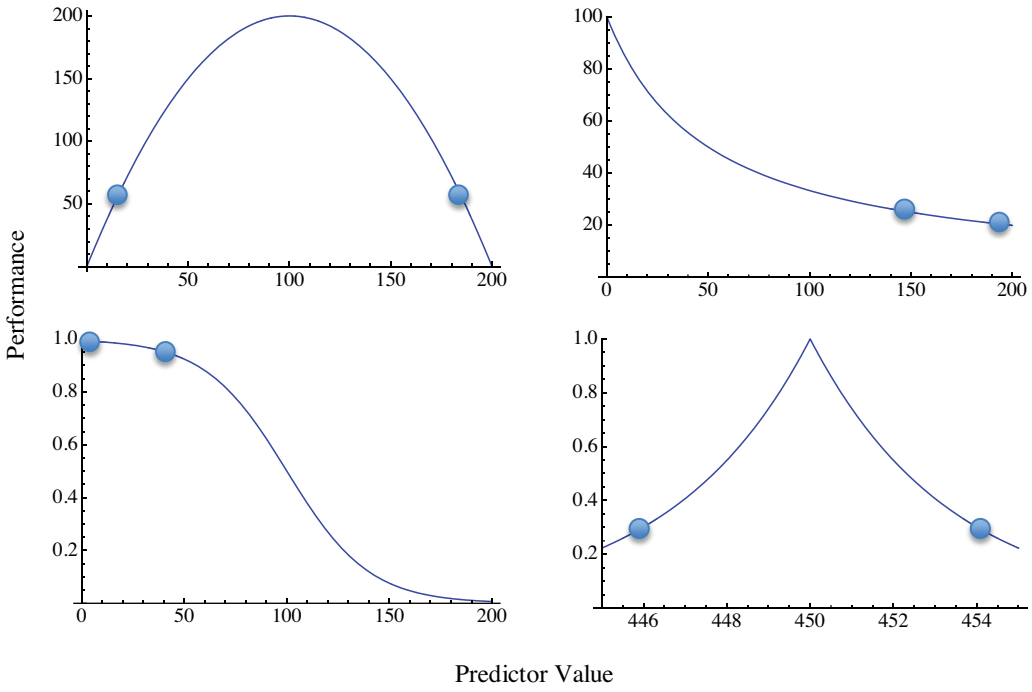

Fig. 1. Four examples of nonlinear relationships commonly observed in psychology: quadratic (upper left), hyperbolic (upper right), logistic (lower left), and generalization (lower right). Each graph shows two points along these continua that can generate similar or identical outcomes.

Many important findings are being missed or misrepresented due to failures to use the best available methodologies (e.g., Bates and Watts, 2007; Bolker et al., 2008; Erceg-Hurn and Mirosevich, 2008; Fox, 2000; Hastie and Tibshirani, 1990; McCullagh and Nelder, 1989; Pinheiro and Bates, 2004). Inevitably, these failures retard scientific progress. Although there are multiple reasons for this slow rate of adoption that include inertia and insufficient training, my focus concerns the problem of *categorical thinking*.

What do I mean by categorical thinking? In this manuscript, I will examine four issues that I believe reflect a categorical approach to psychology that is limiting progress: (a) only examining two or three points along a continuum, (b) turning continuous data into categorical data through median splits and their variations, (c) treating graduated variables like trial and block as categorical in an analysis, and (d) the constraints inherent in being tied too closely to analysis of variance (ANOVA). It is important to be mindful that not all researchers demonstrate one or more of these categorical approaches to their research data, that nearly all of us have used one or more of these techniques on occasion, and that in some cases this categorical thinking is purposeful and reflects a particular research goal. The problems arise when these approaches become entrenched and thus retard progress in a field of investigation. There are other domains in which concerns about categorical thinking have been discussed (e.g., the emphasis on significance rather than effect size, and the problems with categorical diagnosis in the Diagnosis and Statistical Manual, Cohen, 1994; Trull and Durrett, 2005), but these have received significant coverage elsewhere and are less related to the theme underlying the four issues that I will discuss.

Learning to think more continuously about our designs and data presents significant challenges, from designing experiments in order to explore continua to the use of new graphing techniques and statistical analysis. The problem is amplified by the skepticism of reviewers when submitted articles use these unfamiliar methods and, if accepted, the barriers that readers will encounter if the presentation of the research lacks clarity and assumes too much of the reader. It is imperative that an author make an adequate effort to ensure that any approach that is new to the audience be adequately explained and justified. But, it is also critical that we equip ourselves and the next generation of researchers with the tools to achieve success as we move into this foreign territory.

The goal of this paper is not to be exhaustive nor to be accusatory but merely to improve practice by assembling these topics in an accessible location and presenting recommendations for change. I will focus my discussion on those issues that are most relevant to the field of comparative psychology and draw on examples from that field and my own research. However, these issues are of broad relevance to the field of psychology.

2. Examples of categorical thinking that limits progress

2.1. Limited sampling of a continuum

A quick glance at the experimental psychology literature reveals that most researchers sample only two or three levels of a continuum when designing their experiments (two or three delays, doses, or reward magnitudes) and when more than three levels are sampled, the vast majority of the time these values are treated as categorical in an ANOVA rather than as points along a continuum in a regression-based approach. The advantages of only sampling two or three levels of a predictor include the simplicity of the design, the ability to use ANOVA to analyze the data, and the ease of presentation of the results. The principal drawback is that we gain little insight into the functional relationship between the predictors and the outcome (for a philosophical discussion of this issue in psychology, see Lewin, 1931, 1936).

When we sample only two points along a continuum, an infinite number of functions can connect these two points and thus we cannot discern the nature of the relationship. Indeed, even the absence of a difference in the outcome across these two predictor values must be treated with caution because the relationship could be nonlinear and the two points sampled may inadvertently generate similar responses. For example, relationships between a predictor and an outcome like those shown in Fig. 1 reveal that