Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Organizational Behavior and Human Decision Processes 99 (2006) 202–211
ORGANIZATIONAL BEHAVIOR AND HUMAN DECISION PROCESSES
www.elsevier.com/locate/obhdp

Differences between probability and frequency judgments: The role of individual differences in working memory capacity ☆

Amber Sprenger *, Michael R. Dougherty

Department of Psychology, University of Maryland, College Park, MD 20782, USA

Received 31 August 2004
Available online 19 September 2005

Abstract

Most theories of probability judgment assume that judgments are made by comparing the strength of a focal hypothesis relative to the strength of alternative hypotheses. In contrast, research suggests that frequency judgments are assessed using a non-comparative process; the strength of the focal hypothesis is assessed without comparing it to the strength of alternative hypotheses. We tested this distinction between probability and frequency judgments using the alternative outcomes paradigm (Windschitl, Young, & Jenson, 2002). Assuming that judgments of probability (but not judgments of frequency) entail comparing the focal hypothesis with alternative hypotheses, we hypothesized that probability judgments would be sensitive to the distribution of the alternative hypotheses and would be negatively correlated with individual differences in working memory (WM) capacity. In contrast, frequency judgments should be unrelated to the distribution of the alternatives and uncorrelated with WM-capacity. Results supported the hypotheses.

Keywords: Probability judgment; Frequency judgment; Support theory; Working memory

Introduction

Consider the following two types of judgments. A doctor stands before file cabinets that contain all of her patients' records for her entire career and estimates how many of the patients died from cancer. Or, the doctor estimates the likelihood that a randomly chosen file would be of a patient who died from cancer. Normatively, these two types of judgments should have a high degree of correspondence, because the probability of pulling a file of a person who died from cancer should reflect the frequencies of patients who died from cancer versus patients who did not die from cancer. Despite this normative isomorphism, research and theory suggests that there are fundamental differences in the processes underlying judgments of probability and frequency. Moreover, these differences in process have often been purported to underlie the common finding that frequency formats lead to improved accuracy on Bayesian inference problems (Cosmides & Tooby, 1996; Fiedler, 1988; Hoffrage, Gigerenzer, & Krauss, 2002; Gigerenzer & Hoffrage, 1995; but see Sloman, Over, Slovak, & Stibel, 2003).

The purpose of our research was twofold. Our primary purpose was to investigate differences in the cognitive processes used to make probability judgments versus frequency judgments. Although theories of probability judgment assume that focal hypotheses[1] are com-

☆ This material is based on work supported by the National Science Foundation under Grant SES-0134678. Order of authorship is arbitrary. The authors thank Steven Pipe for help in data collection and Paul Windschitl for providing us with the experimental stimuli.
* Corresponding author.
E-mail address: asprenger@psyc.umd.edu (A. Sprenger).

[1] Throughout the introduction, we will use the term focal hypothesis to indicate the hypothesis being considered for judgment.

0749-5978/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.obhdp.2005.08.002

pared with alternative hypotheses and theories of frequency judgment do not assume such a comparison process, few direct tests of differences in comparative versus non-comparative judgment processes in probability and frequency judgment have been published. The second purpose of our research was to investigate the accuracy of frequency and probability judgments. As mentioned above, considerable research supports the idea that frequency judgments often are more accurate than probability judgments. Interestingly, however, most studies comparing frequency and probability judgments have relied almost exclusively on what might be termed *absolute accuracy*–measures that essentially amount to an overconfidence measure (i.e., the degree to which judgments exceed an objective standard). Much less research has examined differences in *relative accuracy*, where relative accuracy is defined as one's ability to discriminate between events that have different objective probabilities. It is important to examine both measures of accuracy, because one's definition of accuracy can be pivotal in determining whether or not one kind of judgment is more accurate than the other (Treadwell & Nelson, 1996).

The present research differs from prior investigations of frequency and probability judgments in two important ways. First, most conclusions that frequency judgments are more accurate than probability judgments have been based on experiments using word-based problems (e.g., Tversky & Kahneman's, 1983 Linda problem) or general knowledge questions. In contrast, in the present study we used a learning-based paradigm where participants learned the objective frequencies of the to-be-judged events. A second major difference between our investigation and previous ones is that we examined two aspects of judgment accuracy: absolute accuracy and relative accuracy. In contrast, most studies comparing probability judgment to frequency judgment have relied exclusively on only one measure of accuracy—absolute accuracy (Cosmides & Tooby, 1996; Fiedler, 1988; Gigerenzer & Hoffrage, 1995; Hoffrage et al., 2002; but see Sloman et al., 2003). Thus, claims that frequency judgments are more accurate than probability judgments have been based on only one definition of accuracy.

Comparative versus non-comparative judgment

One of the primary differences between probability and frequency judgment theories is that probability judgments assume that participants compare the strength of the focal event (e.g., cancer) with the strength of a set of alternative events (e.g., no cancer) (Dougherty, Gettys, & Ogden, 1999; Tversky & Koehler, 1994). In contrast, theories of frequency judgment (Hintzman, 1988; Shiffrin, 2003; Murdock, Smith, & Bai, 2001) assume that participants assess a strength or familiarity dimension of only the focal event (e.g., cancer) and map that feeling onto a frequency scale. Thus, the crucial difference between models of probability and frequency judgment is that probability judgments include a comparison of the focal event with the alternative events.

Support theory (Tversky & Koehler, 1994) provides a general theoretical framework for describing the process of comparing alternative events assumed for probability judgment:

$$P(\mathrm{A},\mathrm{B}) = \frac{s(\mathrm{A})}{s(\mathrm{A}) + s(\mathrm{B})}, \qquad (1)$$

where $P(\mathrm{A},\mathrm{B})$ represents the probability of hypothesis A rather than hypothesis B, $s(\mathrm{A})$ represents the support for the focal hypothesis A, and $s(\mathrm{B})$ represents the support for the alternative hypotheses entertained by the decision maker. Tversky and Koehler (1994) noted that people's judgments tend to be subadditive; the probability of an implicit disjunction tends to be lower than the sum of the probabilities assigned to its elements. For example, if one were to judge *p*(cancer, no cancer) it would be judged as less likely than the sum of the probabilities assigned to *p*(lung cancer, no lung cancer), *p*(breast cancer, no breast cancer), *p*(skin cancer, no skin cancer), and *p*(all other cancer). Thus, the judged probability of the inclusive hypothesis, *p*(cancer, no cancer) is subadditive with respect to the sum of the judged probabilities of its elements.

One way to conceptualize the assessment of support of the alternative hypothesis, $s(\mathrm{B})$, within support theory is in terms of the generation of the alternative hypotheses contained within B. Consider Eqs. (2) and (3). Let h represent the focal hypothesis, A, and $\neg h_i$ represent the *i*th element of the set of B alternative hypotheses. Ideally, people should consider the support for all relevant alternatives when making probability judgments:

$$P(h) = \frac{s(h)}{s(h) + \sum_{i=1}^{N} s(\neg h_i)}, \qquad (2)$$

where N is the total number of alternative hypotheses contained in B. Without assuming that participants underestimate the alternatives, $\neg h_i$, Eq. (2) predicts that subjective judgments should be additive. However, if people are limited in the number of alternatives that they can consider at any one point in time, the overall support for the alternatives will be underestimated and the probability of the focal will be overestimated. Dougherty and Hunter (2003a, 2003b) found that the judged probability of a focal hypothesis decreased as the number of alternative hypotheses people generated increased. Thus, we propose that probability estimation is constrained by the number of alternative hypotheses one can compare at one time, as represented in Eq. (3):