*Japanese Psychological Research*
1997, Volume 39, No. 2, 124–129

*Short Report*

# Upward versus downward anchoring in frequency judgments of social facts

KIMIHIKO YAMAGISHI[1]
*Department of Psychology, University of Washington, PO Box 351525, Seattle, WA 98195, USA*

Abstract: Frequency estimation of social facts in two methods of judgment elicitation was investigated. In the "narrow-range" condition, subjects answered questions in the format: "Out of 100 incidents, how many belong to category X?" In the "wide-range" condition, the frequency for the same event was assessed with respect to "Out of 10,000." Judged frequencies in the wide-range condition were divided by 100, and were compared with the corresponding judgments in the narrow-range condition. Such comparisons were made for low-frequency and high-frequency events. Previous research has shown that, for low-frequency events, judged frequencies are *proportionally* greater in the narrow-range than in the wide-range condition. These results reflect cognitive processes of implicit anchoring, whereby judged frequencies lie close to small numbers within the response ranges provided. I call this process "downward anchoring," and predicted that this tendency would be replicated in the present study. Moreover, I predicted that assessments about high-frequency events would evoke similar cognitive processes operating in the opposite direction. By such "upward anchoring," judged frequencies would lie close to relatively larger numbers within the given response ranges. Consequently, I predicted that judged frequencies for high-frequency events would be *proportionally* greater in the wide-range condition than in the narrow-range condition. These predictions were confirmed.

Key words: frequency estimation, probability judgment, anchoring and adjustment, gender stereotype.

People often have to estimate the frequency with which they encounter various kinds of event. Deciding whether you should join an airline company's frequent-flyer plan may call for an estimation of how frequently you would use the company's flights. However, research has shown that such assessments may be consistent yet biased simultaneously.

How are frequency assessments biased but consistent at the same time? Fischhoff, Slovic, and Lichtenstein (1981) reported that subjective frequency estimation of deaths due to life-threatening events was highly correlated with actual frequencies. However, the judged frequencies were too low for high-frequency events and too high for low-frequency events. Yamagishi (1994a, 1994b) noted that frequency estimation of deaths varied as a function of a particular range of response provided in the question. For 11 well known causes of death, Yamagishi's subjects in one condition assessed the frequencies with respect to 100 people in

[1] I would like to thank Richard Gonzalez, John Miyamoto, Karl Halvor Teigen, and participants of the Judgment and Decision Making Seminar at the University of Washington for helpful discussions. Correspondence concerning this article should be addressed to the author, who is now at the School of Business and Environment, Shukutoku University, Fujikubo 1150-1, Miyoshi-cho, Iruma-gun, Saitama 354, Japan. E-mail: kimihiko@shukutoku.ac.jp.

the population (e.g., "Out of 100 people, how many die of cancer?"). In another condition, subjects estimated the frequency for the same set of events in terms of "Out of 10,000 people." The frequency estimates were reliably correlated with the actual occurrence. Yet, the estimates showed systematic "response-range effects" between the "narrow-range" condition (out of 100) and the "wide-range" condition (out of 10,000). The estimates from the wide-range condition were divided by 100 and converted into percentages. For the 11 causes of death, the judged percentages were greater in the narrow range (cancer example: 24.14%) than in the wide range condition (12.86%). Moreover, Yamagishi (1994b) observed that individual subjects provided highly consistent rank-ordering relationships among the 11 causes between the two methods of judgment elicitation. Hence, he remarked that "Probably it is in this relative sense that risk perception can be meaningfully measured" (Yamagishi 1994b, p. 663).

Yamagishi interpreted the effects as reflecting cognitive processes of spontaneous and implicit anchoring. "Anchoring" and "adjustment" were defined as follows:

> … people make estimates by starting from an initial value that is adjusted to yield the final answer. The initial value … may be suggested by the formulation of the problem, or it may be the result of a partial computation. In either case, adjustments are typically insufficient.
>
> (Tversky & Kahneman, 1974, p. 1128)

For example, in estimating a rare frequency (i.e., deaths due to pneumonia), subjects would spontaneously anchor to a relatively small number $X$ in the context of a particular response range (either out of 100 or 10,000). It may be argued that estimated frequencies were greater when they were adjusted from $\frac{X}{100}$ than $\frac{X}{10{,}000}$. Moreover, an insufficient adjustment with respect to 100 would produce more proportional bias than an adjustment with respect to 10,000. Thus, the estimates of infrequent events were low in the narrow condition (Yamagishi, 1994a, 1994b).

However, Yamagishi did not investigate another prediction that would follow from this argument. In Yamagishi's studies, frequencies were always assessed for rare events, such as deaths due to influenza. Hereafter, I refer to events that occur less than 15% of the time as "low-frequency" events. In frequency estimation for low-frequency events, implicit anchors would be set in a relatively lower vicinity of possible response ranges because of the nature of the events being uncommon. I refer to such implicit anchoring as "downward anchoring." Thus, Yamagishi (1994a, 1994b) showed that downward anchoring systematically occurs in frequency estimation of low-frequency events.

This paper investigates a logical implication of the above argument to supplement Yamagishi's (1994a, 1994b) analysis. In doing so, I examined the following conditions. I refer to events that occur more than 85% of the time as "high-frequency" events. In assessments of high-frequency events, implicit anchors would be set in higher vicinity among possible response ranges. I call it "upward anchoring" when larger numbers within a particular response range would serve as an implicit anchor, due to the nature of the event being prevalent. If upward anchoring were in effect in assessments of high-frequency events, a reverse pattern of response-range effects as noted by Yamagishi (1994a) would be observed, because an anchor would be set to relatively large numbers $Y$ and $Y'$ in particular contexts of response ranges (e.g., $\frac{Y}{100}$ or $\frac{Y'}{10{,}000}$). Again, insufficient adjustments with respect to 100 would produce more serious proportional bias than to 10,000, because judged percentages would be lower in the narrow condition. Moreover, regarding social events that are neither low- nor high-frequency phenomena (hereafter referred to as "medium-frequency" events), one would anchor to a center value in a particular response range. Hence, judged frequencies would produce comparable percentages (e.g., $\frac{50}{100}$ or $\frac{5{,}000}{10{,}000}$) between the narrow- and wide-response ranges.

The following analysis tests this set of predictions: Frequency estimations, converted into percentages, would be greater in the "Out of 100" format than in the "Out of 10,000" format when frequencies of low-frequency events are estimated. This trend would reverse when frequencies of high-frequency events are estimated. Finally, regarding medium-frequency events, such response-range effects would cease to occur.

## Method

### Subjects

Subjects were 524 undergraduates at the University of Washington. They were enrolled in an introductory psychology course and participated to earn course credits.

### Material

I used the following criteria to categorize an event as low, medium, or high frequency. Low-frequency events were those that, in the United States (US Department of Commerce, 1993), are less frequent than 15%. Medium-frequency events were those that fall between 40% and 60%. High-frequency events were those that are more frequent than 85%.

I chose the proportion of females or males as an example of such events. As a matter of fact, it is possible to sample certain occupations that are either predominantly male or predominantly female (e.g., 94.3% of registered nurses are female). Three "female" occupations were sampled: librarians (87.6% female), kindergarten teachers (98.6% female), and registered nurses. I also sampled three "male" occupations: clergy, dentists, and engineers (male percentages are 91.6, 91.5, and 91.5, respectively). Finally, I sampled three neutral occupations that qualify as medium-frequency events for females and males. These were auditors, actresses and actors, and book authors (female percentages are 51.2, 54.5, and 40.3, respectively). I remark here that the gender labels bear no sexist intentions, and the use of the quotation marks reflects this idea.

Each subject was required to produce frequency estimates for either males or females, and for either the narrow (out of 100) or wide (out of 10,000) response range. Hereafter I use NF, WF, NM, and WM to denote the response conditions of females in the narrow range, females in the wide range, males in the narrow range, and males in the wide range, respectively. Each subject was presented with a particular occupation with a particular response condition. For instance, one subject was assigned to the NF condition for nurses and to the WM condition for dentists, whereas another was assigned to the NM condition for nurses and to the WF condition for dentists. Each particular configuration of the occupation and the response condition was organized randomly, along with other filler tasks, in a questionnaire booklet. Each subject was presented with all the occupations and the presentation orders were counterbalanced.

### Design

I used a between-subject design regarding each occupation. In the WF and WM conditions, subjects estimated the frequency of females or males, respectively, with respect to 10,000 observations in the population. In the NF and NM conditions, the female/male frequencies were estimated with respect to 100 observations in the population.

### Procedure

Data were gathered in a group setting. Each subject received a questionnaire booklet, and was required to work individually and quietly. They were instructed to estimate the frequencies "according to your intuition."

### Predictions

The following predictions deal with the responses in the WF and WM conditions divided by 100 to convert into percentages. This procedure makes the responses from the WF and WM conditions directly comparable to the NF and NM conditions. Hereafter, I treat all responses as percentages. The three lines below illustrate my predictions for judged frequencies as converted to percentages. To understand this, recall that female frequencies are high-frequency events whereas male frequencies are low-frequency events in the "female" occupations, while female frequencies are low-frequency events and male frequencies are high-frequency events in the "male" occupations.

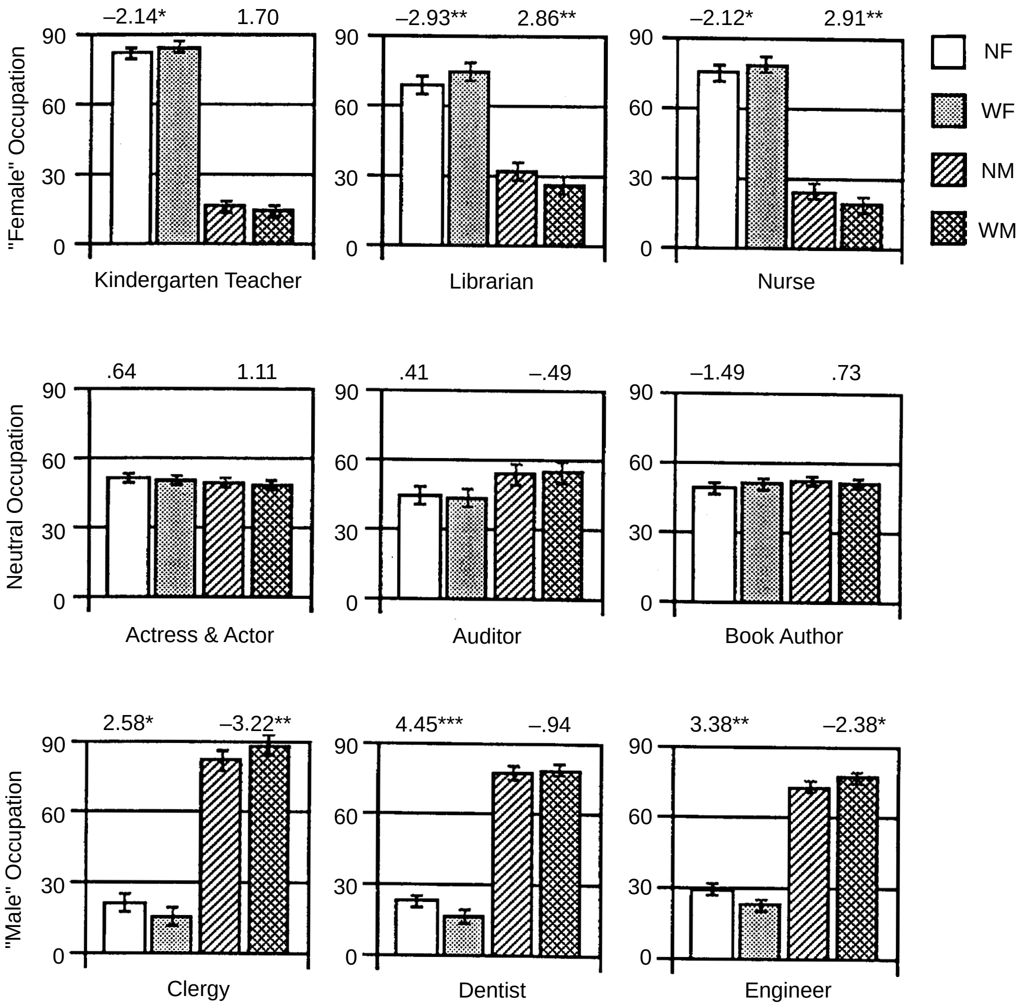


**Figure 1.** Mean judged percentages for the "female," neutral, and "male" occupations.

Finally, both females and males are medium-frequency events in the neutral occupations.

| | |
|---|---|
| "Female" occupations: | NF < WF, NM > WM |
| Neutral occupations: | NF ≈ WF, NM ≈ WM |
| "Male" occupations: | NF > WF, NM < WM |

## Results

The statistical tests reported below are all two tailed. In Figure 1, *, **, and *** denote a significant difference at the .05, .01, and .001 level, respectively.

For each occupation, I performed a 2 × 2 ANOVA (narrow versus wide response range by females versus males) to calculate the *MSE*. Because my goal is to test the differences in specific pairs of group means (as illustrated above), I do not report the results for omnibus *F* tests, which are less informative than planned comparisons. The *MSE* for actor and actress, auditor, author, clergy, dentist, engineer, kindergarten teacher, librarian, and nurse were 71.68, 295.31, 90.18, 279.77, 130.30, 215.30, 93.46, 279.30, and 194.07, respectively. Each *MSE* term has 520 degrees of freedom.

Using the *MSE*, I tested two orthogonal planned comparisons for each occupation. A [1 – 1 0 0] contrast, wherein the elements correspond to the NF, WF, NM, and WM conditions respectively, tested the proportional difference between the NF and WF conditions. The *t*-test statistic would be positive whenever NF > WF and negative whenever NF < WF. A second contrast, [0 0 1 – 1], tested the proportional difference between the NM and WM conditions. The *t*-test statistic would be positive whenever NM > WM and negative whenever NM < WM.

Figure 1 shows mean judged percentages for the NF, WF, NM, and WM conditions for each occupation. For each condition, the "I" bar shows the 95% confidence intervals for the cell mean. As predicted, for the "female" occupations NF < WF and NM > WM. Furthermore, the "male" occupations show that NF > WF and NM < WM. The numbers on the top of each graph shows the *t*-test statistic. For each occupation, the left number shows the *t*-value for the [1 – 1 0 0] contrast and the right number shows the *t*-value for the [0 0 1 – 1] contrast. The predictions were confirmed with statistical significance for the librarian, nurse, clergy, and engineer categories. For kindergarten teacher and dentist, the results of the [1 – 1 0 0] contrasts were significant whereas the [0 0 1 – 1] contrasts failed to reach significance. Yet, the signs of the *t*-values were in the predicted direction. Importantly, none of the neutral occupations showed significant response-range effects, as predicted.

## Discussion

The results mostly supported the prediction that upward anchoring would occur when one estimates frequencies of high-frequency events, whereas downward anchoring would occur when one estimates frequencies of low-frequency events. Concretely, regarding the "female" occupations, the judged frequencies of females were proportionally greater in the NF condition than the WF condition, whereas the judged frequencies of males were proportionally greater in the WM than in the NM condition. In contrast, regarding the "male" occupations, the opposite trend was observed. All the statistically significant differences indicated this set of predicted differences. Finally, regarding the neutral occupations, no response-range effect was found for females and males.

The judgmental task investigated in this paper, namely to estimate relative frequencies, could be regarded as equivalent to judgments of subjective probability. Gigerenzer, Hell, and Blank (1988) proposed two distinct conceptualizations of intuitive probability judgment. A judge is said to be in a "frequency" mode when she/he is assessing the frequency with which a particular class of events occurs. In judging about a single instance, a judge is said to be in a "single-case" mode. Although Gigerenzer et al. remarked that either mode may lead to systematic biases in judgment, they suggested that judgments in the frequency mode tend to be more logically coherent and to adhere to applicable normative constraints. However, Jones, Jones, and Frisch (1995) showed that certain biases in probability judgments are more prevalent in frequency mode. Moreover, the current results indicate that upward and downward anchoring and subsequent biases occur in frequency judgments. Thus, this paper may be characterized as documenting another example that judgment in the frequency mode is nonetheless systematically biased.

## References

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1981). Lay foibles and expert fables in judgments about risks. In T. O'Riordan & R. K. Turner (Eds.), *Progress in resource management and environmental planning* (pp. 161–202). New York: Wiley.

Gigerenzer, G., Hell, W., & Blank, H. (1988). Presentation and content: The use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance*, **14**, 513–525.

Jones, S. K., Jones, K. T., & Frisch, D. (1995). Biases of probability assessment: A comparison of frequency and single-case judgments. *Organizational Behavior and Human Decision Processes*, **61**, 109–122.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, **185**, 1124–1131.

US Department of Commerce (1993). *Statistical abstract of the United States 1992*. Washington, DC: US Government Printing Office.

Yamagishi, K. (1994a). Consistencies and biases in risk perception: I. Anchoring process and response-range effect. *Perceptual and Motor Skills*, **79**, 651–656.

Yamagishi, K. (1994b). Consistencies and biases in risk perception: II. Individual subjects' performance. *Perceptual and Motor Skills*, **79**, 659–663.

(Received April 10, 1995; accepted Sept. 9, 1995)