Measures of Anchoring in Estimation Tasks

Karen E. Jacowitz
University of California, Berkeley

Daniel Kahneman
Princeton University

The authors describe a method for the quantitative study of anchoring effects in estimation tasks. A calibration group provides estimates of a set of uncertain quantities. Subjects in the anchored condition first judge whether a specified number (the anchor) is higher or lower than the true value before estimating each quantity. The anchors are set at predetermined percentiles of the distribution of estimates in the calibration group (15th and 85th percentiles in this study). This procedure permits the transformation of anchored estimates into percentiles in the calibration group, allows pooling of results across problems, and provides a natural measure of the size of the effect. The authors illustrate the method by a demonstration that the initial judgment of the anchor is susceptible to an anchoring-like bias and by an analysis of the relation between anchoring and subjective confidence.

T he terms *anchor* and *anchoring effect* have been used in the psychological literature to cover a bewildering array of diverse experimental manipulations and results, ranging from the effects of unjudged stimuli on psychophysical scales (Helson, 1964) to early bids and offers in negotiations (Neale & Bazerman, 1991). The proliferation of meanings is a serious hindrance to theoretical progress. The present treatment focuses on anchoring effects in tasks of quantitative estimation. In this context, an anchor is an arbitrary value that the subject is caused to consider before making a numerical estimate. An anchoring effect is demonstrated by showing that the estimates of groups shown different anchors tend to remain close to these anchors. We introduce a new technique for the measurement of anchoring effects in this design, and we illustrate the statistical and theoretical analyses that this improved measurement makes possible.

TECHNIQUES AND INTERPRETATIONS

Studies of anchoring in estimation tasks have often used a sequence of two tasks: Subjects first judge whether

a particular value (the anchor) is higher or lower than an uncertain quantity; they then estimate the quantity (e.g., Cervone & Peake, 1986; Plous, 1989; Schkade & Johnson, 1989, Experiment 3; Wright & Anderson, 1989). An effort has sometimes been made to convince subjects that the anchor is arbitrary and uninformative: The selection of the anchor has been variously determined by a wheel of chance (Tversky & Kahneman, 1974), by a randomly chosen card (Cervone & Peake, 1986), or by the experiment number (Switzer & Sniezek, 1991). Wilson, Houston, Etling, and Brekke (1994) used the subject's social security number as an anchor in a task of estimating the number of physicians listed in the phone book. In all these cases, ostensibly uninformative anchors produced large effects.

Perhaps because it is so robust and reliable, the anchoring effect has often been used as a primitive concept, which explains other results but is not itself explained. We have encountered three main ideas about the role of the anchor in tasks of quantitative judgment. These ideas are not mutually exclusive; one or more might be true in different cases or even in the same case.

Starting-point for adjustment. In this model, the subject first considers the anchor, determines whether it is too high or too low, then derives an estimate by adjusting in the appropriate direction until an acceptable value is found. The adjustment is generally insufficient (Tversky & Kahneman, 1974), perhaps because it terminates at

the nearest boundary of a large region of acceptable values (Quattrone, Lawrence, Finkel, & Andrus, 1984; Wilson et al., 1994).

Conversational hint. At least in some experiments, the anchor can be taken as a hint from the experimenter, which subjects clutching at straws quite reasonably use (Kahneman & Tversky, 1982; Schwarz, 1990, 1994; Schwarz, Strack, Hilton, & Naderer, 1991). A number mentioned in a question induces a gradient of plausibility if the subject assumes that the experimenter only asks questions on which opinions are likely to differ.

Suggestion or prime. The intuitive estimation of an uncertain quantity is a complex memory task, which involves implicit, automatic, and uncontrolled activation of candidate answers; the final estimate is often a blend or compromise in which any activated answer is assigned some weight (Strack, 1992; Wilson & Brekke, 1994). The anchor affects the response because it is treated as a candidate answer in this automatic process. This view of anchoring is compatible with Gilbert's (1990) idea that people initially respond to any statement by believing it. Thus a subject who is asked whether the Amazon River is longer or shorter than 5,000 miles may entertain at least a transient belief in both possibilities.

MEASURES OF ANCHORING

Anchoring effects are generally believed to be large and reliable. But how large is large? The present article introduces a yardstick for the measurement of anchoring effects in estimation tasks. The procedure for measuring anchoring requires three groups of subjects drawn from the same population. A calibration group provides estimates of a set of uncertain quantities, without any anchors. Subjects in two other groups offer estimates after judging an anchor. The anchors are selected by their position in the distribution of estimates in the calibration group. In the present experiment, the low and high anchors were respectively set at the 15th and 85th percentiles of estimates for each question.

For descriptive analyses of the anchoring effect, we use an anchoring index (AI) to measure the movement of the median estimate of anchored subjects toward the anchor to which they have been exposed. The AI for a particular estimation problem is defined as follows:

$$AI = \frac{\text{Median (high anchor)} - \text{Median (low anchor)}}{\text{High anchor} - \text{Low anchor}}$$

An AI can also be defined for each anchor separately. For example, the AI for a low anchor is computed as follows:

$$AI(low) = \frac{\text{Median (low anchor)} - \text{Median (calibration group)}}{\text{Low anchor} - \text{Median (calibration group)}}$$

The index for the high anchor is defined similarly. Plausible values of AI range from 0 (no anchoring effect) to 1 (median estimates of anchored subjects coincide with the anchors that they have been shown). Larger values are also possible.

The AI is most useful as a descriptive statistic, which provides a readily interpretable measure of an anchoring effect. Other measures are more appropriate for the purposes of detailed statistical analyses and hypothesis testing. For these purposes, we transform all estimates into the corresponding percentiles in the calibration group. For example, an anchored estimate equal to the median of the calibration group would be assigned a transformed score of 50. Anchored estimates that are outside the range of the responses of the calibration group are assigned values of 100 or 0. The anchoring effect is now measured by comparing these transformed scores in the two anchored groups; a statistical test can be performed using either the Mann-Whitney test or the Student's t. The percentile transformation permits statistical comparisons of anchoring effects on different problems; it also permits meaningful pooling of data over several problems. Of course, the ordinal transformation loses information about the shape of the distribution of judgments.

We report a study of anchoring in an estimation task that illustrates these methods of analysis. Three substantive questions are addressed: (a) What is the size of the anchoring effect in estimates of uncertain quantities? (b) How do the judgments of anchors as high or low relate to the open-ended estimates from which the anchors were derived? and (c) What is the relationship between the size of the anchoring effect and subjects' confidence in their estimates?

METHOD

Subjects were 156 students at the University of California, Berkeley, who completed a questionnaire in partial fulfillment of a course requirement in an introductory psychology class.

Subjects in the calibration group ($n = 53$) were recruited first. They were asked to estimate 15 quantities and to rate their confidence in each answer on a 10-point scale. As shown in Table 1, the quantities to be estimated included the height of Mount Everest and the number of members in the United Nations. The 15th and the 85th percentiles of the distribution of estimates in the calibration group were used as anchors for the experimental (anchored) groups.

Experimental subjects ($n = 103$) answered three consecutive questions about each of the 15 quantities. They first indicated whether the quantity was greater or less than an anchor value; next, they estimated the quantity;

TABLE 1: Summary Statistics for Calibration and Anchored Estimates

Question	Calibration Median	Anchors Low	Anchors High	Medians Low Anchor	Medians High Anchor	Transformed Medians Low Anchor	Transformed Medians High Anchor	Extreme Values (%) Low	Extreme Values (%) High	Anchoring Index
1. Length of Mississippi River (in miles)	800	70	2,000	300	1,500	33.96	76.42	10	37	.62
2. Height of Mount Everest (in feet)	12,000	2,000	45,500	8,000	42,550	36.79	83.02	16	48	.79
3. Amount of meat eaten per year by average American (in pounds)	180	50	1,000	100	500	37.74	80.19	10	16	.42
4. Distance from San Francisco to New York City (in miles)	3,200	1,500	6,000	2,600	4,000	37.74	74.53	6	18	.31
5. Height of tallest redwood (in feet)	200	65	550	100	400	31.13	77.36	12	31	.62
6. Number of United Nations members	50	14	127	26	100	28.30	78.30	20	24	.65
7. Number of female professors at the University of California, Berkeley	50	25	130	50	95	41.51	60.38	18	28	.43
8. Population of Chicago (in millions)	1.0	0.2	5.0	0.6	5.05	33.65	88.46	8	52	.93
9. Year telephone was invented	1889	1850	1920	1870	1900	27.55	64.29	28	8	.43
10. Average number of babies born per day in the United States	2,000	100	50,000	1,000	40,000	37.74	84.91	8	42	.78
11. Maximum speed of house cat (in miles per hour)	18	7	30	10	20	25.00	61.25	18	20	.43
12. Amount of gas used per month by average American (in gallons)	50	20	80	40	60	35.90	74.36	4	30	.33
13. Number of bars in Berkeley, CA	20	10	85	20	40	45.00	68.75	14	12	.27
14. Number of state colleges and universities in California	30	20	100	30	50	55.00	72.50	22	20	.25
15. Number of Lincoln's presidency	16	7	17	16	16	64.10	64.10	10	12	.00

then they indicated their confidence in the estimate. There were two versions of the questionnaire, each with high anchors for some quantities and low anchors for the others. The quantities appeared in the same order in all questionnaires.

RESULTS AND DISCUSSION

Table 1 shows, for each problem, the median of the calibration group and the two anchors; it also shows, for both high and low anchors, the median estimate, the median of transformed estimates (i.e., the percentile score in the calibration group corresponding to the median anchored estimate), and the proportion of estimates more extreme than the anchor. Finally, the last column of Table 1 shows the AI computed for each problem.

The anchoring effects demonstrated in Table 1 are remarkably large: Over the 15 problems, the overall mean of the AI is .49. Thus the median subject moved almost halfway toward the anchor, from the estimate that the subject would have made without it. Another measure of the size of the effect is the correlation between subjects' estimates and the anchor they had seen. We computed the point-biserial correlation in the pooled data of the two anchored groups for each problem; the mean of these correlations over the 15 problems was .42.

Effects of High and Low Anchors

An unexpected observation is that the effects of high and low anchors were not equally strong; the mean AI was .51 for the high anchors and .40 for the low anchors. To test the reliability of this difference, we transformed the anchored estimates into the corresponding percentiles of the calibration group. The median of these scores was 76 for high anchors and 36 for low anchors. The deviation of these values from 50 indicates an anchoring effect, which is larger for high than for low anchors. There was a highly significant difference between individual averages of the transformed estimates made with high and low anchors, $t(102) = 7.99$, $p < .01$. Thus high and low anchors were both very effective, but the effect of high anchors was significantly larger.

Table 1 provides a clue to the effectiveness of high anchors: 27% of estimates made with a high anchor were higher than the anchor, in contrast to 15% such estimates in the calibration group. Thus estimates made with a high anchor differ in two ways from unanchored estimates: (a) Some estimates that would otherwise be low are pulled up toward the anchor and (b) some estimates that would otherwise be lower than the anchor (12% of the total in these data) are pulled up beyond the anchor. Low anchors produced the first of these effects but not the second: Only 14% of estimates made

with a low anchor were lower than the anchor. The difference between the relative frequencies of extreme estimates in the two anchoring conditions was highly significant, $t(102) = 6.12$, $p < .001$.

Anchored estimates were almost always (99.5%) consistent with the just-preceding judgment of the anchor. Thus it is also the case that high anchors were judged low on 27% of occasions, whereas low anchors were judged high only 14% of the time. The order of the judgment and estimation tasks implies that many extreme estimates were mediated by the prior judgment of a high anchor as too low. We are, therefore, led to ask, What is the process by which an anchor is judged too high or too low? The simplest model of this judgment is that the subject first generates an independent estimate of the quantity and then compares the anchor to this estimate. In this simple model, the initial estimate of the quantity is not affected by the anchor; it is presumably the same estimate that the subject would make in response to an open-ended question. Thus we might expect the same percentage of extreme estimates in the open-ended and in the anchored conditions. This expectation was confirmed in the case of low anchors, but it clearly failed with high anchors, which yielded an average of 27% extreme estimates instead of the expected 15%.

The idea that the anchoring effect only occurs after the production of an unbiased estimate is implicit in the models that attribute anchoring to insufficient adjustment (Quattrone et al., 1984; Tversky & Kahneman, 1974; Wilson et al., 1994). The most explicit version of this model (Quattrone et al., 1984) suggested that the adjustment starts at the anchor and ends at the nearest boundary of the region of values that the subject considers possible. This model implies an anchoring bias on the estimate of the quantity (when the anchors are outside the region of uncertainty) but no bias in the judgment of the anchor itself. The observed effects of the high anchors contradict this model, lending support to the alternative hypothesis that (a) the anchor alters the subject's beliefs and (b) the anchor is judged by the altered beliefs. It appears to be the case that the question "Is the true value of Quantity A higher or lower than X?" can, at least when X is high, increase the plausibility of values that are even higher than X. The asymmetric effect of high and low anchors may arise from an asymmetry of uncertainty in many of our problems, in which there is a definite lower bound (zero) but no definite upper bound.

We examined the relation between open-ended estimates and dichotomous judgments of the anchor in other experiments. The first was a close replication of the present study, using most of the same problems, with an added manipulation intended to discredit the informational value of the anchor. The manipulation of credibility did not reduce the anchoring effect, and the pattern of results was the same as in the present study: The high anchors (again located near the 85th percentile of the calibration distribution) were judged low 28% of the time; the low anchors (located at the 15% percentile) were judged high on 15% of occasions.

A subsequent study (Green, Jacowitz, Kahneman, & McFadden, 1995) compared anchoring effects in the estimation of uncertain quantities and in questions about willingness to pay (WTP) for two public goods: "saving about 50,000 seabirds each year from the effects of offshore oil spills" and "achieving a 20% reduction in automobile accidents in the state over 5 years." Data were also obtained for three estimation problems similar to those used in the present study. A calibration group provided unanchored numerical estimates and statements of WTP. Five different anchors were selected from the distribution of these responses. These anchors were used in five separate groups. Subjects in the anchored condition first made a dichotomous judgment ("Is the quantity more than X?") or answered an equivalent referendum question ("Would you be willing to pay X dollars?") and then stated their best estimate or maximal WTP. In accord with the present results, high anchors yielded a large proportion of answers that were even higher than the anchor. For example, 24% of anchored estimates equalled or exceeded the highest anchor in the three estimation problems, but only 4.2% of calibration answers were in that range. Similarly, 15% of the statements of WTP by respondents shown the highest anchor were at least as high as that anchor; only 4.6% of calibration responses were as high. However, the results of that study differed from the present findings in one respect: A low anchor (set at the 25th percentile of calibration responses) also pulled responses up; only about 15% of respondents judged the low anchor high or refused to pay the suggested amount.

The main implication of the results reviewed in this section is that—contrary to common usage—the anchoring effect is not restricted to tasks in which a numerical response is produced. Considering a specific numerical value as a possible answer to a question is evidently sufficient to alter subjects' beliefs, attitudes, or intentions about that question. Thus a question of the form "Is X higher or lower than . . ." can sometimes induce an anchoring-like effect on itself.

Anchoring and Confidence

Several authors have reported that the size of the anchoring effect varies inversely with subjects' confidence in their answers (Quattrone et al., 1984; Wilson et al., 1994). The present study examines the relation between anchoring and confidence by answering several questions.

Is there more anchoring on questions that are generally answered with low confidence? Over the 15 problems listed in Table 1, the correlation between the AI and the mean confidence in the calibration group was $-.68$ ($p < .05$).

Are the responses that are most strongly affected by an anchor made with relatively low confidence? To answer this question, we transformed the estimates made by the anchored subjects into the corresponding percentile scores in the calibration group, as explained earlier. We then computed for each question, separately for the subjects shown the high and the low anchors, the correlation between these transformed estimates and confidence. If susceptibility to anchoring is associated with low confidence, the correlation between estimates and confidence ratings should be negative when the anchor is high and positive when the anchor is low. Over the 15 problems, the mean correlations were $-.14$ for high anchors, $t(14) = 2.37$, $p < .05$, and $.27$ for low anchors, $t(14) = 4.80$, $p < .001$.

Are confident subjects less susceptible to anchoring because they have superior knowledge? We split each of the anchored groups by confidence, separately for each question. We then computed the median estimate offered by each of these groups and checked which of the medians was closer to the true value of the quantity. The confident subjects were closer to the truth on 8 of 15 problems and further from the truth on 5 problems; the distances were equal for the remaining 2 problems. This weak result makes it unlikely that the correlation between confidence and susceptibility to anchoring was mediated by degree of knowledge.

Are highly confident subjects immune to anchoring? For each problem, we selected the 25% of estimates that were associated with the highest confidence ratings, separately for the groups presented with high and with low anchors. We then computed an AI using the medians of these two highly confident groups. The mean value of the AI over the 15 problems was .28, clearly less than the value of .49 observed for the group as a whole but still quite substantial.

What is the effect of presenting an anchor on confidence levels? Overall, estimates were made with greater confidence in the anchored groups than in the calibration group (the mean confidence ratings were 3.85 and 2.99, respectively), $t(154) = 3.53$, $p < .001$. The subjects evidently treated the anchor as useful information.

GENERAL DISCUSSION

The techniques that we have introduced for the analysis of anchoring are potentially useful in several ways. They permit comparisons of anchoring effects across problems stated in different units. They also permit pooling of data over several problems, increasing the power of statistical analyses. However, two important limitations of our measures should be noted. First, the computation of the AI retains the metric of the original estimates, but the psychological significance of that metric is in doubt when the distribution of responses is highly skewed—as often happens. The problem is mitigated, but not solved, by the use of medians, which are relatively insensitive to extreme measures. Second, we used a transformation of estimates into percentiles of the calibration group to conduct statistical tests. This procedure also reduces the impact of wild guesses, but it involves a nonlinear transformation that could affect the results of parametric statistical tests. As usual, such cautions about the scale are most relevant to the interpretation of weak trends.

In the present study, the new measures provided apparently robust evidence that the question "Is the quantity higher or lower than X?" causes anchoring effects that are (a) surprisingly large, (b) sometimes evident in the original evaluation of the anchor as high or low, as well as in subsequent numerical estimates, and (c) inversely related to subjects' confidence in their judgments but substantial even in judgments made with high confidence. As we have noted, the theoretical investigation of anchoring is still quite rudimentary. The finding that anchoring is not restricted to numerical answers is important to this debate because it suggests that anchoring may occur without an adjustment process. Future advances in the study of anchoring will require both a careful taxonomy of the varieties of anchoring effects (e.g., Wilson & Brekke, 1994) and experimental designs and measures that support precise analysis.

REFERENCES

Cervone, D., & Peake, P. (1986). Anchoring, efficacy, and action: The influence of judgmental heuristics on self-efficacy judgments and behavior. *Journal of Personality and Social Psychology, 50,* 492-501.

Gilbert, D. T. (1990). How mental systems believe. *American Psychologist, 46,* 107-119.

Green, D., Jacowitz, K. E., Kahneman, D., & McFadden, D. (1995). *Referendum contingent valuation, anchoring, and willingness to pay for public goods.* Manuscript submitted for publication.

Helson, H. (1964). *Adaptation level theory: An experimental and systematic approach to behavior.* New York: Harper.

Kahneman, D., & Tversky, A. (1982). On the study of statistical intuitions. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases.* Cambridge: Cambridge University Press.

Neale, M. A., & Bazerman, M. H. (1991). *Cognition and rationality in negotiations.* New York: Free Press.

Plous, S. (1989). Thinking the unthinkable: The effects of anchoring on likelihood estimates of nuclear war. *Journal of Applied Social Psychology, 19,* 67-91.

Quattrone, G. A., Lawrence, C. P., Finkel, S. E., & Andrus, D. C. (1984). *Explorations in anchoring: The effects of prior range, anchor extremity, and suggestive hints.* Unpublished manuscript, Stanford University, Stanford, CA.

Schkade, D., & Johnson, E. (1989). Cognitive processes in preference reversals. *Organizational Behavior and Human Decision Process, 44,* 203-231.

Schwarz, N. (1990). Assessing frequency reports of mundane behaviors: Contributions of cognitive psychology to questionnaire construction. In C. Hendrick & M. S. Clark (Eds.), *Review of personality and social psychology: Vol. 11. Research methods in personality and social psychology* (pp. 98-119). Newbury Park, CA: Sage.

Schwarz, N. (1994). Judgment in a social context: Biases, shortcomings, and the logic of conversation. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 26, pp. 125-162). San Diego: Academic Press.

Schwarz, N., Strack, F., Hilton, D., & Naderer, G. (1991). Base rates, representativeness, and the logic of conversation: The contextual relevance of "irrelevant" information. *Social Cognition, 9,* 67-84.

Strack, F. (1992). The different routes to social judgments: Experiential versus informational strategies. In L. Martin & A. Tesser (Eds.), *The construction of social judgment* (pp. 249-275). Hillsdale, NJ: Lawrence Erlbaum.

Switzer, F. S., & Sniezek, J. A. (1991). Judgment processes in motivation: Anchoring and adjustment effects on judgment and behavior. *Organizational Behavior and Human Decision Process, 49,* 208-229.

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

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin, 116,* 117-142.

Wilson, T. D., Houston, C., Etling, K. M., & Brekke, N. (1994). *Anchoring versus adjustment errors in judgment: The generalizability of anchoring effects.* Working paper, University of Virginia.

Wright, W., & Anderson, U. (1989). Effects of situation familiarity and financial incentives on use of the anchoring and adjustment heuristic for probability assessment. *Organizational Behavior and Human Decision Process, 44,* 68-82.