ORGANIZATIONAL BEHAVIOR AND HUMAN DECISION PROCESSES 49, 188–207 (1991)

Overconfidence in Estimation: Testing the Anchoring-and-Adjustment Hypothesis

RICHARD A. BLOCK AND DAVID R. HARPER

Montana State University

People are frequently overconfident in the accuracy of their estimates of uncertain quantities. The present study requested 50%- or 90%-confidence ranges. Overconfidence is shown when less than the target percentage of ranges include the true value. Tversky and Kahneman (1974) proposed that people use an anchoring-and-adjustment heuristic: They begin with a starting value, one supplied to them or generated by them, and insufficiently adjust their estimates around this anchor. The present data support the proposed anchoring process. If subjects receive another person's point estimates, their own implicit point estimates are correlated with these values. However, anchoring-and-adjustment processes do not invariably produce overconfidence. Subjects who receive anchors are no more overconfident than are those who do not receive anchors. If subjects are required to produce a point estimate first, overconfidence decreases; processes involved in explicitly displaying the point estimate are implicated. Overconfidence may occur because people do not realistically assess their estimation ability.

A person usually does not obtain all of the relevant information before making a judgment or decision. Even if the information is potentially available, such as in reference books, a person may not invest the time and energy needed to obtain it. Cognitive conceit—that is, overconfidence in personal cognitive attributes, such as knowledge and decision-making abilities—may also decrease the likelihood that someone will seek potentially important information. Instead, a person may simply estimate relevant quantities, such as numerical attributes of objects and probabilities of events. Difficulties may ensue if a person has only imprecise information available in memory concerning these objects or events.

Alpert and Raiffa (1982) conducted an important study in which subjects estimated ten uncertain quantities, such as the total US egg produc-

Address correspondence to Richard A. Block, Department of Psychology, Montana State University, Bozeman, MT 59717-0001. David R. Harper is now at Bowling Green State University. We thank the students who were involved in collecting and analyzing these data, especially Joseph Mangiantini and Win Turner. We also thank Robert Patterson, Wesley Lynch, and several anonymous referees for helpful comments on the manuscript. Some of these data were presented at the 1986 and 1987 meetings of the Psychonomic Society and at the Fourteenth Interdisciplinary Conference in 1989.

0749-5978/91 $3.00

tion in 1965. Alpert and Raiffa explained in detail how the subjects should assess probability distributions for these quantities. They used the method of direct fractile assessment; for example, a person gives a lower quartile estimate (i.e., a .25 fractile) and an upper quartile estimate (i.e., a .75 fractile). These two estimates define a 50%-confidence interval, or a range that the person thinks is as likely to include the true value as to exclude it. If a person estimates several quantities, the percentage of ranges that include the true value can be compared to the target percentage, which in this example is 50%. If someone realistically assesses the extent of his or her knowledge concerning the quantities, in the long run exactly 50% of the ranges should contain the true value. Alpert and Raiffa found, instead, that only about 33% of the ranges did. This is a classic overconfidence effect, since people appear to be unjustifiably confident in the accuracy of their ranges (cf. Adams & Adams, 1961). Pitz (1974) obtained similar data showing what he called a *hyperprecision effect*.[1] He also found that subjects are less overconfident when they estimate quantities about which they presumably have greater knowledge. Lichtenstein, Fischhoff, and Phillips (1982) concluded that "the most pervasive finding in recent research is that people are overconfident with general-knowledge items of moderate or extreme difficulty" (p. 314) and that "overconfidence is most extreme with tasks of great difficulty" (p. 315).

Tversky and Kahneman (1974) suggested that a person performing this kind of task implicitly uses a heuristic called *anchoring and adjustment*. Along with two more well-known heuristics (representativeness and availability), people are assumed to use this heuristic in the process of making judgments under conditions of uncertainty. Tversky and Kahneman reported an anchoring effect in an experiment in which subjects received arbitrary starting values. Each subject estimated various quantities, such as the percentage of African countries belonging to the United Nations. Before each estimate, the experimenter spun a "wheel of fortune" in the subject's presence, and the subject first judged whether the resulting number was higher or lower than the true value of the quantity. Then the subject estimated the quantity by adjusting upward or downward from the starting value. Median estimates of the percentage in

[1] Peterson and Pitz (1988) recently studied how subjects use information in a prediction task. They preferred the term *uncertainty* to the ones used in previous articles, *hyperprecision* (Pitz, 1974) and *overconfidence* (Peterson & Pitz, 1986). In this usage, the term *uncertainty* refers to a belief about possible values of a quantity, and *confidence* refers to a belief that a previously stated prediction is accurate. Thus, greater overconfidence reflects less uncertainty. We prefer the term *confidence*, rather than *uncertainty*, in the context of the present task because of the abundant historical precedents to use the term *overconfidence effect* to describe the phenomenon being investigated.

this example were 25 and 45 for subjects that received anchoring values of 10 and 65, respectively. This demonstrates that a supplied anchoring value influences point estimates. However, it does not directly test the hypothesis that people use anchoring-and-adjustment processes when they are not given an anchor, when they are producing confidence ranges, or both. In addition, this finding may not directly relate to the phenomenon of overconfidence.

Tversky and Kahneman claimed, however, that subjects use an anchoring-and-adjustment heuristic when making direct fractile assessments of subjective probability distributions (as well as other kinds of estimates). They conducted an additional demonstration which used two conditions. In one condition, subjects estimated quantities (e.g., the Dow-Jones average on a future date) by producing .10 and .90 fractiles. These subjects "were too extreme. . . . Events that they defined as having a probability of .10 actually obtained in 24 percent of the cases" (Tversky & Kahneman, 1974, p. 1130). We can calculate that these subjects' mean 80%-confidence range (the interval between their .10 and .90 fractiles) included the true value in only 52% of the cases. Thus, these subjects showed a classic overconfidence effect.

In another condition, subjects were given each median judgment by subjects in the first condition. The experimenter asked them to judge the odds that each of these values exceeded the true value of the quantity. These subjects "were too conservative. Events to which they assigned an average probability of .34 actually obtained in 26 percent of the cases" (Tversky & Kahneman, 1974, p. 1130). We can calculate that these subjects' mean 32%-confidence range (the interval between their .34 and .66 fractiles) would have included the true value in 48% of the cases. A natural anchor (i.e., a .50-probability judgment) apparently influenced subjects, and an underconfidence effect resulted. Tversky and Kahneman concluded that "the degree of calibration [i.e., confidence effect] depends on the procedure of elicitation" (p. 1130).

Tversky and Kahneman also said that "subjects state overly narrow confidence intervals which reflect more certainty than is justified by their knowledge . . . This effect is attributable, in part at least, to anchoring" (p. 1129). If there is a natural starting value for the estimation process or if an experimenter provides a starting value, this self-generated or externally supplied anchoring value presumably influences the person. If the person must provide a point estimate, it may be closer to the anchoring value than it would be if no anchor were generated or supplied. If a person must produce a confidence range, he or she may insufficiently adjust the limits of this range away from the anchoring value. Tversky and Kahneman did not discuss effects of supplied anchors on confidence ranges; however, they suggested that this anchoring-and-adjustment process occurs even if someone is not supplied an anchoring value and is not required to produce a point estimate.

Consider a hypothetical example. Here are the thoughts of a person attempting to produce a 98%-confidence range concerning the number of foreign cars imported into the United States in 1968:

> I think there were about 180 million people in the U.S. in 1968; there is about one car for every three people thus there would have been about 60 million cars; the lifetime of a car is about 10 years, this suggests that there should be about 6 million new cars in a year but since the population and the number of cars is increasing let's make that 9 million for 1968; foreign cars make up about 10% of the U.S. market, thus there were probably about 900,000 foreign imports. . . .

The person has now generated a point estimate, or anchoring value. This leads the person insufficiently to spread the lower and upper estimates:

> To set my 98% confidence band, I'll add and subtract a few hundred thousand cars from my estimate of 900,000 (Slovic, Kunreuther, & White, 1974, p. 195).

Thus, the person produces a confidence range from about 600,000 to about 1,200,000, an insufficiently wide range.

Most researchers ignore the finding that underconfidence may also be a result of anchoring processes and discuss only how anchoring-and-adjustment leads to the more typical finding of overconfidence in estimation. Researchers have interpreted various kinds of overconfidence in terms of anchoring and adjustment (e.g., Davis, Hoch, & Easton Ragsdale, 1986; Einhorn & Hogarth, 1985; Lopes, 1987; Lovie, 1985; Slovic, Fischhoff, & Lichtenstein, 1977). Anchoring-and-adjustment explanations also appear widely in books on organizational behavior and management science (e.g., Bazerman, 1986), cognitive psychology (e.g., Matlin, 1989), and even introductory psychology (e.g., Bootzin, Bower, Zajonc, & Hall, 1986).

The present study provides the first direct test of an anchoring-and-adjustment explanation for overconfidence and thereby confronts the issue of what judgment and decision processes this kind of estimation task involves. Experiment 1 replicated typical findings regarding two major variables that influence overconfidence: whether or not subjects have personal familiarity with or knowledge about the quantities and whether or not the experimenter warns subjects not to be overconfident in the accuracy of their estimates. Experiments 2–6 directly tested the anchoring-and-adjustment hypothesis to determine whether it, or some alternative explanation, characterizes the processes involved in producing confidence ranges. To our knowledge, no researchers have directly tested whether anchoring-and-adjustment processes produce overconfidence (or underconfidence) in a situation in which a person does not receive anchoring values, but in which it is assumed that the person generates anchors nevertheless.

Although most researchers have been primarily concerned with 90%-, 95%-, and 99%-confidence ranges (a notable exception being Murphy & Winkler, 1977), for several reasons Experiments 1–4 used 50%-confidence ranges. First, the concept of a 50%-confidence range is relatively easy to explain, and the participants (few of whom presumably had much statistical training) seemed to easily grasp the concept. Second, using 50%-confidence ranges helps avoid potential floor and ceiling effects. Finally, with 50% confidence as a target there is considerable sensitivity to find potential underconfidence (in which more than 50% of the ranges include the true value). In Experiments 5 and 6, 90%-confidence ranges were used to extend the generality of the findings.

EXPERIMENT 1

Because this relatively simple confidence-range task had not been used much, we started by replicating two major findings. Several of our experiments, including Experiment 1, replicated the finding that subjects are less overconfident when they estimate familiar quantities than when they estimate unfamiliar quantities (e.g., Pitz, 1974). One can view this independent variable in other ways, however. For example, people presumably find unfamiliar quantities more difficult to estimate than familiar quantities. If subjects realistically assess their knowledge, they should produce wider confidence ranges for unfamiliar items than for familiar ones. If subjects do increase their ranges, an additional question is whether they will increase them enough to compensate for their relative lack of knowledge about unfamiliar quantities compared to their relative abundance of experience concerning familiar quantities. In a preliminary experiment, we found overconfidence in 50%-confidence ranges on unfamiliar quantities, but no significant overconfidence on familiar quantities.

In another preliminary experiment, we found that simply informing subjects about the overconfidence effect did not decrease it. In Experiment 1, some subjects were not only informed about the effect, but also warned not to be overconfident. This is an example of a debiasing procedure (for reviews, see Fischhoff, 1982; Winterfeldt & Edwards, 1986), which also include giving subjects extensive training with feedback (e.g., Alpert & Raiffa, 1982; Lichtenstein & Fischhoff, 1980; Sharp, Cutler, & Penrod, 1988) and giving subjects an algorithm to structure their estimates (e.g., MacGregor, Lichtenstein, & Slovic, 1988). Alpert and Raiffa (1982) gave some subjects extensive feedback describing typical overly confident performance. Subjects were also told: "For heaven's sake, *Spread Those Extreme Fractiles!* Be honest with yourselves! Admit what you don't know!" (p. 301). With this admonition, the percentage of 50%-confidence ranges that included the true value increased from 33 to 43%. So overconfidence decreased, but it was not eliminated.

Method

Materials and procedure. Each subject was asked to produce a 50%-confidence range for each quantity. We explained that this is "a range of possible values" such that for each item the person "believes the range is as likely to include the true value as it is to not include the true value." Thus, each subject provided both a lower estimate and an upper estimate. (Note that .25 and .75 fractiles were not explicitly requested; the instructions were simpler than those used in most previous experiments.) Subjects received a page of instructions. It contained four sample estimates, two of which included the true value and two of which did not, and a brief commentary on each.

Uninformed subjects received no additional information. Warned subjects, however, also read the following, which was printed in capital letters on a separate page: "When people make estimates like these, they are usually too confident that their estimates are accurate. . . . Be sure not to be overly confident in your knowledge."

Half of the subjects estimated 12 familiar quantities, and the other half estimated 12 unfamiliar quantities. Familiar quantities are those that subjects presumably had perceptually experienced. For example, subjects estimated the number of spokes on a single wheel of an adult bicycle, as well as the height of a standard soft-drink can (in inches). Unfamiliar quantities are those that subjects could not have (or probably never had) perceptually experienced. For example, subjects estimated the area (in square miles) of the island of Singapore, as well as the number of US states that voted to ratify the Equal Rights Amendment. To eliminate any possible scaling bias, the true value of each unfamiliar item was closely matched with that of a familiar item. True values ranged from 4.0 to 228.9, and across items the logarithms of these values were spaced fairly evenly from .60 to 2.36.

Subjects. A total of 100 introductory psychology students volunteered for the experiment in exchange for course credit. We randomly assigned 25 participants to each combination of conditions in a 2 × 2 (Information × Item Type) factorial design. Subjects participated in one of several sessions, and approximately the same number of subjects in a session served in each condition.

Results

Table 1 shows means on three measures, each of which provides an informative index of performance (cf. Peterson & Pitz, 1986).

TABLE 1
MEANS ON CONFIDENCE-RANGE MEASURES IN EXPERIMENT 1

Condition	n	In-range percentage[a]	Range ratio	Error ratio
		Familiar quantities		
Uninformed	25	49.3 ± 4.2	1.7	3.3
Warned	25	54.3 ± 2.9	1.8	3.6
		Unfamiliar quantities		
Uninformed	25	28.7 ± 3.5	3.6	6.5
Warned	25	38.3 ± 4.1	4.4	6.5

[a] Each mean in-range percentage is shown with its standard error.

In-range percentage. The *in-range percentage* is the percentage of ranges that included the true value. A 2 × 2 analysis reveals that the in-range percentage was greater for familiar than for unfamiliar quantities [$F(1,96) = 24.4$, $p < .001$]. Warned subjects had a greater in-range percentage than did uninformed subjects [$F(1,96) = 3.93$, $p < .05$]. The two variables did not interact ($F < 1$).

Overconfidence is revealed if the mean in-range percentage is significantly less than 50%. Subjects who estimated familiar quantities were not significantly overconfident (or underconfident) in either the uninformed or the warned condition [$t(24) = .16$ and 1.48, respectively]. Subjects who estimated unfamiliar quantities were overconfident in both the uninformed [$t(24) = 6.14$, $p < .001$] and the warned condition [$t(24) = 2.88$, $p < .01$].

Confidence range. Subjects tended to produce lower and upper estimates in a way based on ratios of numbers (see later). Thus, the geometric mean (across items and subjects) of the ratio of the upper estimate to the lower estimate provides a measure of the width of the confidence range.[2] The middle column of Table 1 shows this measure, hereafter called the *range ratio.*

Item type influenced subjects' confidence ranges [$F(1,96) = 82.4$, $p < .001$]. On familiar quantities, the mean upper estimate is only 1.8 times the mean lower estimate; on unfamiliar quantities, the mean upper estimate is 4.0 times the mean lower estimate. Information condition did not significantly influence confidence ranges [$F(1,96) = 2.47$], and item type and information condition did not interact significantly ($F < 1$).

[2] A few subjects gave an occasional lower estimate of 0. To avoid an undefined ratio in these cases, we first added 1 to all estimates. (This is a common practice when using logarithmic transformations.)

Estimation error. To measure the absolute error in subjects' estimates, we first calculated what we call the *implicit point estimate* of each quantity. It is the geometric mean of the lower and upper estimates (after 1 is added to each value; see Footnote 2). Each implicit point estimate was then compared to the actual value of the quantity (plus 1) by taking the ratio of the higher value to the lower value. This so-called *error ratio* is shown in the right-hand column of Table 1.

Item type influenced estimation error [$F(1,96) = 121.6$, $p < .001$]. Estimates of familiar quantities are in error by a factor of 3.5, whereas estimates of unfamiliar quantities are in error by a factor of 6.5. Neither the effect of information condition nor its interaction with item type is significant (both $F < 1$).

Discussion

Experiment 1 replicated some classical findings on overconfidence. Subjects are less overconfident on familiar than on unfamiliar quantities (e.g., Pitz, 1974). On familiar quantities, subjects produced narrow confidence ranges and (at least in Experiment 1) were not significantly overconfident. On unfamiliar quantities, subjects produced wider confidence ranges, so they apparently were somewhat aware of their limited knowledge about the quantities. However, they did not increase their ranges enough to compensate for their limited knowledge. Experiment 1 also replicated the finding that warned subjects are less overconfident, but they are still overconfident on unfamiliar items (e.g., Alpert & Raiffa, 1982).

EXPERIMENTS 2–4

Experiment 2 directly tested the anchoring-and-adjustment hypothesis. If subjects begin the estimation process by implicitly generating a point estimate, even though the instructions do not request one, then requiring subjects to begin by generating a point estimate should not greatly influence their overconfidence. In other words, a strict version of the anchoring-and-adjustment hypothesis says that subjects who are not required to provide a point estimate should be as overconfident as subjects who are required to do so. Although Tversky and Kahneman (1974) did not discuss this issue, if an explicit point estimate is required, it may produce a more salient anchoring effect; hence, overconfidence may be somewhat greater than in a no-anchor condition. The data of the present Experiment 2 revealed, instead, that subjects are less overconfident in a self-anchor than in a no-anchor condition.

Experiment 3 tested two possible reasons for this decreased overconfidence in the self-anchor condition. One possibility is that the mere presence of an explicit point estimate is responsible. Another is that subjects' awareness of the processes (or of the results of processes) involved in generating a point estimate is responsible. To test between these possibilities, subjects were paired: Each person in an other-anchor condition received the point estimates of a self-anchor subject. If the mere presence of anchoring values influences subjects to decrease their overconfidence, the two conditions should not differ in overconfidence. If, instead, processes involved in generating a point estimate are necessary, self-anchor subjects should be less overconfident than other-anchor subjects. (The other-anchor condition is also interesting in that it reveals something about judgment processes in a social context.) The data of Experiment 3 revealed less overconfidence in the self-anchor than in the other-anchor condition.

Experiment 4 replicated various findings of Experiments 2 and 3 in a single experiment so that additional comparisons could be made. In particular, if the presence of a salient anchor invariably leads to insufficient adjustment of lower and upper estimates, then other-anchor subjects should be more overconfident than no-anchor subjects.[3] The data of Experiment 4 rejected this notion.

Method

Materials and procedure. Experiments 2–4 all used similar materials. Item type (i.e., familiarity) was manipulated in the same way as in Experiment 1. The other variable, anchoring condition, involved a change in instructions and judgment format. For no-anchor subjects, the instructions and judgment sheet were like those in Experiment 1. Self-anchor subjects were instructed to write a point estimate of each quantity before giving the lower and upper values defining the 50%-confidence range. Each other-anchor subject received the point estimates that a self-anchor subject had generated. Instructions said that these estimates were another person's estimates of the quantities, and that the subject was free to consider or to ignore them when producing the lower and upper values.

Subjects. In Experiment 2, 76 volunteers were recruited from an introductory psychology class, and 19 of them were randomly assigned to each of four combinations of conditions. The 2 × 2 (Item Type × Anchoring) factorial design used unfamiliar versus familiar items and no-anchor versus self-anchor conditions.

In Experiment 3, 46 subjects were recruited from the university com-

[3] Independently of any proposed anchoring-and-adjustment process, other-anchor subjects may display more overconfidence than no-anchor subjects if they consider the other person's estimate to be credible information that they combine with their own estimate in a Bayesian way. The present data reveal, however, no significant difference in overconfidence (see General Discussion).

munity and were paired. One subject of each pair was randomly assigned to the self-anchor condition, and one was assigned to the other-anchor condition. All subjects estimated unfamiliar items.

In Experiment 4, 120 volunteers were recruited from an introductory psychology class, and 20 of them were randomly assigned to each of six combinations of conditions. The 2 × 3 (Item Type × Anchoring) factorial design used familiar versus unfamiliar items and no-anchor, self-anchor, and other-anchor conditions.

Results

Table 2 shows means on the three measures in each combination of conditions.

In-range percentage. In Experiments 2 and 4, the mean in-range percentage is greater for familiar than unfamiliar quantities [$F(1,72) = 11.7$ and $F(1,114) = 15.2$, both $p < .005$]. All three experiments show an effect of anchoring condition [$F(1,72) = 6.42$, $t(22) = 3.63$, and $F(2,114) = 3.11$, respectively, all $p < .05$]. The interaction of item type and anchoring condition is not significant in Experiments 2 or 4 (both $F < 1.15$).

Three comparisons, combining results across experiments (whenever

TABLE 2

MEANS ON CONFIDENCE-RANGE MEASURES IN EXPERIMENTS 2–4

Condition	Experiment	n	In-range percentage[a]	Range ratio	Error ratio
		Familiar quantities			
No anchor	2	19	49.6 ± 4.0	1.8	3.5
	4	20	47.5 ± 3.9	1.9	3.5
Self anchor	2	19	58.3 ± 4.6	2.0	3.4
	4	20	52.9 ± 3.9	1.9	3.6
Other anchor	4	20	50.8 ± 4.6	1.7	3.5
		Unfamiliar quantities			
No anchor	2	19	34.6 ± 3.1	3.3	5.7
	4	20	35.0 ± 3.8	4.0	6.9
Self anchor	2	19	46.1 ± 3.7	3.9	6.3
	3	23	41.7 ± 3.7	4.2	5.5
	4	20	46.7 ± 4.1	4.3	5.5
Other anchor	3	23	26.8 ± 2.0	2.6	5.5
	4	20	32.9 ± 2.4	2.8	5.6

[a] Each mean in-range percentage is shown with its standard error.

possible), clarify the effect of anchoring condition.[4] The in-range percentage is greater for self-anchor than no-anchor ($Z = 3.16$, $p < .001$) and other-anchor conditions ($Z = 2.74$, $p < .01$), which do not differ significantly [$t(78) = 0.15$].

On familiar quantities, subjects are not significantly overconfident (or underconfident) in any of the anchoring conditions [$Z = .53$, $Z = 1.74$, and $t(19) = .18$, respectively]. On unfamiliar quantities, estimates of no-anchor, self-anchor and other-anchor subjects are all overconfident ($Z =$ 5.06, 2.35, and 6.88, all $p < .02$).

Confidence range. As in Experiment 1, subjects in Experiments 2 and 4 produced wider confidence ranges on unfamiliar than on familiar quantities [$F(1,72) = 99.2$ and $F(1,114) = 106.5$, both $p < .001$]. Anchoring condition did not influence confidence ranges in Experiment 2 [$F(1,72) =$ 3.35], but it did in Experiments 3 and 4, which both included an other-anchor condition [$t(22) = 3.83$ and $F(2,114) = 6.55$, both $p < .005$]. Familiarity and anchoring condition did not interact significantly in Experiments 2 or 4 (both $F < 1.62$).

Several comparisons clarify the effect of anchoring condition. Confidence ranges of no-anchor and self-anchor subjects did not differ significantly ($Z = .42$). Other-anchor subjects, however, produced narrower ranges than both no-anchor [$t(78) = 2.09$, $p < .05$] and self-anchor subjects ($Z = 3.27$, $p < .005$).

Estimation error. As before, absolute estimation error calculations used implicit point estimates.[5] In Experiments 2 and 4, subjects estimated familiar items more accurately than unfamiliar items [$F(1,72) = 75.7$ and $F(1,114) = 114.9$, both $p < .001$]. Anchoring condition did not significantly influence estimation error in any of the experiments [$F < 1$, $t(22)$ $= 1.06$, and $F(2,114) = 2.04$, respectively]. Familiarity and anchoring condition did not interact significantly in Experiments 2 or 4 (both $F <$ 2.23).

Additional analyses. Self-anchor subjects' data allow us to assess their interpretation of instructions and their production of lower and upper values. Assume that when subjects give a 50%-confidence range, they produce approximately the .25 and .75 fractiles and that they do so in a way based on ratios of numbers rather than in a way based on differences between numbers. If this is the case, self-anchor subjects should produce

[4] The method of adding weighted Zs was used to combine results across experiments (see Rosenthal, 1978).

[5] If the actual point estimates made by self-anchor subjects are used, the mean error ratio on each item type is not appreciatively different, no significant effects become nonsignificant, and no nonsignificant effects become significant.

a mean ratio of the point estimate to the lower value that is approximately equal to the ratio of the upper value to the point estimate. Averaged across experiments, on familiar items the mean ratio of the point estimate to the lower value is 1.4 and that of the upper value to the point estimate is also 1.4. These ratios are not significantly different ($Z = 1.08$). On unfamiliar items, the mean ratio of the point estimate to the lower value is 2.1 and that of the upper value to the point estimate is 2.0. These ratios are also not significantly different ($Z = .62$).

Several correlational analyses reveal that supplied anchors influenced other-anchor subjects' implicit point estimates.[6] For other-anchor subjects, the mean correlation between the log of the supplied anchoring value and the log of the implicit point estimate is .40 on unfamiliar and .38 on familiar quantities. In contrast, for no-anchor subjects the correlation between logs of randomly-paired, but unsupplied, values and logs of implicit point estimates is close to zero, .02 on unfamiliar and −.01 on familiar quantities.

Discussion

Experiments 2–4 reveal several findings. Most importantly, requiring subjects to provide explicit point estimates decreases overconfidence. (This is the case even though neither their confidence ranges nor their estimation accuracy is significantly greater than that of no-anchor subjects. Apparently the combination of a slight increase in confidence range and a slight increase in accuracy was enough for the self-anchor subjects' in-range percentage to increase relative to that of no-anchor subjects.) This finding does not support a strict interpretation of the anchoring-and-adjustment hypothesis, which says that self-anchor subjects should be at least as overconfident as no-anchor subjects.

Self-anchor subjects were also less overconfident than other-anchor subjects, a finding that implicates processes involved in generating and displaying the point estimate, not merely the presence of an already-generated anchoring value. Compared to other-anchor subjects, self-anchor subjects produced wider confidence ranges, but their implicit point estimates were no more accurate.

Other-anchor subjects were no more overconfident than no-anchor subjects. In Experiment 4, other-anchor subjects did produce slightly narrower confidence ranges than did no-anchor subjects; however, the results of Experiment 6 did not replicate this finding (see later). Although the nonsignificant in-range percentage difference was in the predicted

[6] These correlations were calculated for each item, then averaged across items and, finally, across experiments.

direction in Experiment 4, the results of Experiment 6 also showed no significant difference between these two conditions. This finding rejects the notion that anchoring-and-adjustment processes invariably produce overconfidence.

Ancillary data are consistent with the notion that subjects produce lower and upper estimates in a way based on ratios of numbers. Subjects did not simply add and subtract a constant number from their point estimate as the quotation cited earlier implies (Slovic *et al.*, 1974). In addition, other-anchor subjects' implicit point estimates were closer to the supplied anchoring values than would be expected by chance, so supplied anchors influenced implicit point estimates.

EXPERIMENTS 5 AND 6

Experiment 5 attempted to clarify the processes responsible for the reduced overconfidence of self-anchor subjects compared to no-anchor subjects. Experiments 2–4 had revealed that this decreased overconfidence cannot be attributed to the mere presence of a point estimate—the source of the point estimate matters. Experiment 5 tested whether processes involved in generating point estimates are responsible or whether processes involved in explicitly displaying point estimates on the judgment sheet are responsible. To do this, we used an implicit-anchor condition in which subjects were told to make a point estimate before performing the confidence-range task, but in which they were not instructed to record the point estimate on the judgment sheet. The results reveal that the performance of implicit-anchor subjects does not differ from that of no-anchor subjects. A possible explanation is that no-anchor subjects generate an implicit point estimate at the start of the estimation process, just as the anchoring-and-adjustment hypothesis suggests. The results also showed that overconfidence in self-anchor subjects does not decrease simply because they implicitly generate point estimates. Only if subjects also explicitly display their point estimates do they show less overconfidence.

Experiment 6 tested the notion that in the process of generating and displaying an explicit point estimate subjects realize how poorly they are able accurately to estimate the quantity, and so they widen their confidence ranges. Some subjects rated their ability to accurately estimate each quantity before they generated a confidence range. The ability rating followed one of three anchoring conditions: a no-anchor condition, an explicit-anchor condition, and an other-anchor condition.

In contrast to Experiments 1–4, in Experiments 5 and 6 all subjects were asked to produce 90%-confidence ranges. We switched to this range partly to ascertain the generality of our findings (that is, to determine whether earlier findings would be replicated with a different confidence range). In addition, research on calibration reveals that people are more overconfident at higher confidence levels, such as 90%, than at lower confidence levels, such as 50% (Lichtenstein *et al.*, 1982).

Method

Materials and procedure. Experiments 5 and 6 used four anchoring conditions. For no-anchor subjects, the instructions and materials were like those given to no-anchor subjects in Experiments 2 and 4. For explicit-anchor subjects, the instructions and materials were like those previously given to self-anchor subjects. Implicit-anchor subjects, like explicit-anchor subjects, were told to "start by guessing the actual value of the answer." In contrast to explicit-anchor subjects, implicit anchor subjects were told merely to "keep this best-guess estimate in mind as you adjust it to write your low estimate and your high estimate." As in Experiments 3 and 4, the judgment sheets of each other-anchor subject displayed the point estimates that an explicit-anchor subject had generated, and these subjects were so-informed. For all subjects, the examples and instructions described 90%-confidence ranges, and 10 unfamiliar quantities were estimated.

Experiment 6 used two other conditions, a rating condition and a no-rating condition. In the rating condition, immediately above each line requesting a 90%-confidence range was a line reading, "I would rate my ability to accurately estimate this quantity as . . .", followed by a 10-point scale in which 1 was labeled *very poor* and 10 was labeled *very good*. This rating came after explicit-anchor subjects had generated a point estimate and after other-anchor subjects had read the supplied point estimate; for no-anchor subjects, it was the first task. In the no-rating condition, as in all previous experiments, no ability rating was requested.

Subjects. In Experiment 5, there were 120 participants, some recruited from an introductory psychology class and some from the university community. We randomly assigned 40 participants to each of three anchoring conditions, the no-anchor, implicit-anchor, and explicit-anchor conditions. Data of an additional 3 subjects in the implicit-anchor condition were discarded because they wrote point estimates on the judgment sheet.

In Experiment 6, 150 volunteers were recruited from introductory and advanced psychology classes at two universities. Then 25 students were randomly assigned to each combination of conditions in a 3 × 2 (Anchoring × Ability Rating) factorial design that used no-anchor, explicit-anchor, and other-anchor conditions, along with rating versus no-rating conditions.

Results

Table 3 shows means on the three measures in each combination of conditions.

In-range percentage. Anchoring condition influenced subjects' in-range percentage in both experiments [$F(2,117) = 3.29$ and $F(2,144) = 4.72$, both $p < .05$]. In Experiment 6, neither the effect of rating condition nor its interaction with anchoring condition significantly affected the in-range percentage (both $F < 1.18$).

Five comparisons, combining results across both experiments (whenever applicable) and across rating conditions (in Experiment 6) clarify the effect of anchoring condition. The in-range percentage is greater for explicit-anchor subjects than for both no-anchor ($Z = 3.19$, $p < .001$) and other-anchor subjects [$t(98) = 2.94$, $p < .005$], which do not differ significantly [$t(98) = .47$]. The in-range percentage of implicit-anchor and no-anchor subjects is not significantly different [$t(78) = .13$]. Explicit-anchor subjects show a greater in-range percentage than implicit-anchor subjects [$t(78) = 2.12$, $p < .05$].

All four conditions display overconfidence (all $t > 14.2$, $p < .001$). It is interesting that the overall mean in-range percentage is not substantially different from what it was in Experiments 1–4, which requested a 50%- rather than a 90%-confidence range.

Confidence range. In Experiment 5, anchoring condition influenced confidence ranges [$F(2,117) = 3.60$, $p < .05$]. In Experiment 6, the main effect of anchoring condition is not significant ($F < 1$), but rating condition influenced confidence ranges [$F(1,144) = 4.31$, $p < .05$]; the inter-

TABLE 3
MEANS ON CONFIDENCE-RANGE MEASURES IN EXPERIMENTS 5–6

Condition	Experiment	n	In-range percentage[a]	Range ratio	Error ratio
No anchor	5	40	34.5 ± 2.8	3.9	5.4
	6	25	34.8 ± 3.4	3.5	5.6
	6	25[b]	34.4 ± 3.5	4.7	6.2
Implicit anchor	5	40	35.0 ± 2.6	3.7	5.5
Explicit anchor	5	40	43.8 ± 3.2	5.6	5.0
	6	25	45.2 ± 2.9	4.2	5.4
	6	25[b]	39.6 ± 3.1	4.6	5.6
Other anchor	6	25	30.8 ± 3.0	3.3	5.1
	6	25[b]	35.2 ± 3.6	4.7	5.5

[a] Each mean in-range percentage is shown with its standard error.
[b] Subjects who provided an ability rating.

action of anchoring condition and rating condition is not significant ($F < 1$).

Combining across experiments (whenever applicable) reveals the following pattern of findings. Explicit-anchor subjects produced wider ranges than did both no-anchor ($Z = 3.19$, $p < .001$) and implicit-anchor [$t(78) = 2.26$, $p < .05$] subjects. There is no significant difference between the explicit-anchor and other-anchor conditions [$t(98) = .68$], the no-anchor and other-anchor conditions [$t(98) = .17$], and the implicit-anchor and no-anchor conditions [$t(78) = .29$].

Estimation error. Anchoring condition did not significantly influence estimation error (see Footnote 5) in either experiment (both $F < 1$). In Experiment 6, the estimation-error data show neither a significant effect of rating condition nor a significant interaction of rating condition and anchoring condition (both $F < 1$).

Additional analysis. Averaged across explicit-anchor subjects in both experiments, the mean ratio of the point estimate to the lower value is 2.2 and that of the upper value to the point estimate is 2.1. These two ratios are not significantly different ($Z = 1.04$). The results of this analysis, like those in Experiments 2–4, are consistent with the notion that subjects produce estimates in a way based on ratios of numbers.

As before, correlational analyses reveal that supplied anchors influenced other-anchor subjects' point estimates (see Footnote 6). Across other-anchor subjects in Experiment 6, the mean correlation between the log of each supplied anchoring value and the log of the implicit point estimate is .53. In contrast, the mean correlation between logs of randomly paired, but unsupplied, values and logs of implicit point estimates is $-.01$ for no-anchor and .03 for implicit-anchor subjects.

Ability rating. In Experiment 6, the mean ability rating is 4.30 in the no-anchor condition, 3.52 in the explicit-anchor condition, and 5.12 in the other-anchor condition. Anchoring condition influenced this rating [$F(2,72) = 6.72$, $p < .005$]. A Newman–Keuls test clarifies this effect. Compared to no-anchor subjects, explicit-anchor subjects were less confident in their ability accurately to estimate the actual values of the quantities ($p < .05$). Also compared to no-anchor subjects, other-anchor subjects were more confident in their ability ($p < .01$). Other-anchor subjects (who had the same anchors, but who obtained them differently) were also more confident in their ability than were explicit-anchor subjects ($p < .01$).

Discussion

Experiments 5 and 6 replicate and extend certain findings of Experiments 2–4: (1) Requiring subjects to provide explicit point estimates decreases their overconfidence compared to a no-anchor condition. (2) Explicit-anchor subjects are also less overconfident than are other-anchor subjects. (3) Supplied anchors influence implicit point estimates, but other-anchor subjects are no more overconfident than are no-anchor subjects. (4) The overconfidence of implicit-anchor subjects is about the same as that of no-anchor subjects. (5) Compared to implicit-anchor subjects, explicit-anchor subjects show wider confidence ranges and less overconfidence. Taken together, this pattern of results reveals that overconfidence decreases only if a person must explicitly generate and display a point estimate before producing the confidence range. No-anchor subjects, like implicit-anchor subjects, may generate an implicit point estimate at the start, but this alone is not sufficient to reduce overconfidence.

As in Experiments 2–4, in-range percentage differences between anchoring conditions cannot be attributed to any significant difference in accuracy of implicit point estimates. The decreased overconfidence of explicit-anchor subjects compared to all other subjects is reflected simply in their production of wider confidence ranges.

In Experiment 6, subjects who rated their ability accurately to estimate the quantities produced wider confidence ranges than did those who did not rate their ability. A possible explanation is suggested in the General Discussion.

GENERAL DISCUSSION

Experiment 1 replicated previous findings regarding overconfidence in estimation. When people estimate unfamiliar quantities, their confidence ranges are too narrow, a classic overconfidence effect. Performance on familiar quantities is better: Confidence ranges are narrower, and yet overconfidence decreases. If people receive a salient warning that they should not be too confident in their knowledge, overconfidence on unfamiliar items decreases, although it is not eliminated.

Results of Experiments 2–6 support the notion of anchoring. Supplying a person with someone else's point estimates influences the person's point estimates. However, the mere presence of an anchoring value does not invariably produce insufficient adjustment of estimates and overconfidence. Combining data across experiments, confidence ranges were no more narrow for other-anchor than for no-anchor subjects ($Z = 1.41$, $p = .16$), and other-anchor subjects were no more overconfident than were no-anchor subjects.

Tversky and Kahneman (1974), as well as Slovic and his colleagues (Slovic *et al.*, 1974; Slovic & Lichtenstein, 1971), claimed that people implicitly generate and use an anchoring value at the start of an estimation process (such as producing a confidence range), even if the task does not explicitly require one. If this is the case, asking subjects explicitly to provide a point estimate before giving a confidence range should not influence (or should actually slightly increase) their overconfidence compared to subjects who are not asked to provide a point estimate. In contrast, the present data show that requiring subjects to provide an explicit point estimate reduces their overconfidence. These findings suggest that anchoring-and-adjustment processes do not inevitably operate in the simple way previously thought to produce overconfidence in estimation. Explicit self-anchoring decreases overconfidence; it does not produce it.

The mere presence of an anchoring value at the start of an estimation process is not the critical factor underlying this reduced overconfidence. A person who receives someone else's point estimates is more overconfident than is a person who must generate and explicitly display his or her own point estimates. However, simply requiring a person to generate an anchoring value at the start of an estimation process is not sufficient. Instead, subjects must generate and explicitly display an anchoring value.

Experiment 6 interposed an ability rating between the hypothesized anchoring and adjustment stages to clarify a possible way in which anchoring may influence adjustments. When a person receives someone else's point estimates, he or she reports an increased ability to estimate the quantities accurately. Under these conditions, the person does not realistically assess his or her ability, as revealed by the relatively high ability ratings and yet considerable overconfidence. However, if a person must generate and explicitly display a point estimate, confidence in estimation ability decreases markedly. The difficulty of having to provide this estimate may make the person realize that his or her estimation ability is limited, as revealed by the relatively low ability ratings and less overconfident estimates.

These findings have implications for improving the ability of estimators or forecasters to give accurate confidence ranges. Someone who is estimating a familiar quantity will show little or no overconfidence (or underconfidence), at least for moderate (e.g., 50%) confidence ranges. Because overconfidence in estimating unfamiliar quantities is especially pervasive at higher confidence levels (e.g., 90%), even experts may need to take steps to decrease such inappropriate confidence. Other research suggests that improved forecasting results from such relatively time-consuming methods as: thinking of reasons why a point estimate might be inaccurate (Hoch, 1985; Koriat, Lichtenstein, & Fischhoff, 1980), giving people relatively easy practice items followed by discouraging feedback (Arkes, Christensen, Lai, & Blumer, 1987, Experiment 1), and leading people to expect a group discussion of their estimates (Arkes *et al.*, 1987, Experiment 2). It is nevertheless true that many efforts to decrease overconfidence have failed to do so (see Fischhoff, 1982). The present study suggests that a relatively simple method can be effective: Require a person explicitly to display a point estimate before giving a confidence range.

However, if the situation calls for a high confidence level (e.g., 90%), a person may still be overconfident. This method may be more practical if it is used along with other debiasing techniques.

REFERENCES

Adams, J. K., & Adams, P. A. (1961). Realism of confidence judgments. *Psychological Review,* **68,** 33–45.

Alpert, M., & Raiffa, H. (1982). A progress report on the training of probability assessors. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 294–305). Cambridge, England: Cambridge Univ. Press.

Arkes, H. R., Christensen, C., Lai, C., & Blumer, C. (1987). Two methods of reducing overconfidence. *Organizational Behavior and Human Decision Processes,* **39,** 133–144.

Bazerman, M. H. (1986). *Judgment in managerial decision making.* New York: Wiley.

Bootzin, R. R., Bower, G. H., Zajonc, R. B., & Hall, E. (1986). *Psychology today: An introduction.* New York: Random House.

Davis, H. L., Hoch, S. J., & Easton Ragsdale, E. K. (1986). An anchoring and adjustment model of spousal predictions. *Journal of Consumer Research,* **13,** 25–37.

Einhorn, H. J., & Hogarth, R. M. (1985). Ambiguity and uncertainty in probabilistic inference. *Psychological Review,* **92,** 433–461.

Fischhoff, B. (1982). Debiasing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 422–444). Cambridge, England: Cambridge Univ. Press.

Hoch, S. J. (1985). Counterfactual reasoning and accuracy in predicting personal events. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **11,** 719–731.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory,* **6,** 107–118.

Lichtenstein, S., & Fischhoff, B. (1980). Training for calibration. *Organizational Behavior and Human Performance,* **26,** 149–171.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 305–334). Cambridge, England: Cambridge Univ. Press.

Lopes, L. L. (1987). Procedural debiasing. *Acta Psychologica,* **64,** 167–185.

Lovie, P. (1985). A note on an unexpected anchoring bias in intuitive statistical inference. *Cognition,* **21,** 69–72.

Matlin, M. (1989). *Cognition* (2nd ed.) New York: Holt, Rinehart & Winston.

MacGregor, D., Lichtenstein, S., & Slovic, P. (1988). Structuring knowledge retrieval: An analysis of decomposed quantitative judgments. *Organizational Behavior and Human Decision Processes,* **42,** 303–323.

Murphy, A. H., & Winkler, R. L. (1977). Reliability of subjective probability forecasts of precipitation and temperature. *Applied Statistics,* **26,** 41–47.

Peterson, D. K., & Pitz, G. F. (1986). Effects of amount of information on predictions of uncertain quantities. *Acta Psychologica,* **61,** 229–241.

Peterson, D. K., & Pitz, G. F. (1988). Confidence, uncertainty, and the use of information. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **14,** 85–92.

Pitz, G. F. (1974). Subjective probability distributions for imperfectly known quantities. In L. W. Gregg (Ed.), *Knowledge and cognition* (pp. 29–41). Potomac, MD: Erlbaum.

Rosenthal, R. (1978). Combining results of independent studies. *Psychological Bulletin,* **85,** 185–193.

Sharp, G. L., Cutler, B. L., & Penrod, S. D. (1988). Performance feedback improves the resolution of confidence judgments. *Organizational Behavior and Human Decision Processes*, **42**, 271–283.

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1977). Behavioral decision theory. *Annual Review of Psychology*, **28**, 1–39.

Slovic, P., Kunreuther, H., & White, G. F. (1974). Decision processes, rationality, and adjustment to natural hazards. In G. F. White (Ed.), *Natural hazards: Local, national, global* (pp. 187–205). New York: Oxford Univ. Press.

Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance* **6**, 649–744.

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

von Winterfeldt, D., & Edwards, W. (1986). *Decision analysis and behavioral research*. Cambridge, England: Cambridge, Univ. Press.

RECEIVED: October 31, 1989