Journal of Behavioral Decision Making
J. Behav. Dec. Making, **20**: 179–201 (2007)
Published online 29 September 2006 in Wiley InterScience
(www.interscience.wiley.com) **DOI**: 10.1002/bdm.549

More than, Less than, or Minimum, Maximum: How Upper and Lower Bounds Determine Subjective Interval Estimates

KARL HALVOR TEIGEN[1]*, ANNE-MARIE HALBERG[1]
and KNUT INGE FOSTERVOLD[2]
[1]*University of Oslo, Norway*
[2]*Lillehammer University College, Norway*

ABSTRACT

Uncertain quantities can be described by single-point estimates of lower interval bounds (X_1), upper interval bounds (X_2), two-bound estimates (separate estimates of X_1 and X_2), and by ranges ($X_1 - X_2$). A price estimation task showed that single-bound estimates phrased as "T costs more than X_1" and "T costs less than X_2," yielded much larger intervals than "minimum X_1" and "maximum X_2." This difference can be attributed to exclusive interpretations of X_1 and X_2 in the first case (X_1 and X_2 are unlikely values), and inclusive interpretations in the second (X_1 and X_2 are likely values). This pattern of results was replicated in other domains where participants estimated single targets. When they estimated a distribution of targets, the pattern was reversed. "Minimum" and "maximum" values of variable quantities (e.g., flight prices) were found to delimit larger intervals than "more than" and "less than" estimates. Copyright © 2006 John Wiley & Sons, Ltd.

KEY WORDS interval estimates; uncertainty; inclusive and exclusive intervals; overconfidence

INTRODUCTION

When people are asked to give subjective estimates of a target quantity T, they often do so by suggesting uncertainty intervals. These can be given as complete ranges, but often a single lower or upper bound is suggested: "T costs more than X_1," or "less than X_2." Furthermore, the terms used to delimit the range can be *exclusive* ($X_1 < T$, $X_2 > T$), as in this example, or *inclusive* ($X_1 \leq T$, $X_2 \geq T$), as when we estimate minimum and maximum prices. The studies to be reported in the present paper focus on a previously unexplored topic, namely the effects of inclusive and exclusive terms on single-bound estimates. It turns out that these effects also depend upon the nature of the target to be estimated: T can be a specific object or a distribution of objects.

* Correspondence to: Karl Halvor Teigen, Department of Psychology, University of Oslo, P.O.B. 1094, Blindern, N-0317 Oslo, Norway.
E-mail: k.h.teigen@psykologi.uio.no

The uncertainty interval will in the first case express imperfect knowledge (ignorance), and in the second external variability. A brief introduction to each of these three topics, ranges versus single bounds, inclusive versus exclusive terms, and sources of uncertainty, is given below.

Ranges versus one-sided intervals

Uncertain quantities are often expressed in terms of a confidence interval. I plan to finish a job in 2–4 weeks, or I assume that I will have to pay $1000–1500 for a conference trip. Sometimes only the lower or the upper bounds are estimated: “Global temperatures will increase with at least 2 degrees,” “I will return before midnight.” The degree of confidence associated with such intervals can be indicated by a verbal phrase or by a numeric qualifier, like “I believe that . . . ” or “it is 90% certain that . . . .” For instance, authors submitting manuscripts to *Psychological Science* are told by the editor: “We *hope* to inform you of its status *within 8–12 weeks*” (italics ours). In this case the interval is qualified with a phrase most language users will understand as expressing (a) a polite, positive wish, and (b) a non-trivial, but not very high probability.

The accuracy of such intervals is reflected in their hit rates. How often are editorial decisions actually made within 8–12 weeks? In the calibration literature, hit rates are compared to degree of confidence, expressed as probability statements (Griffin & Brenner, 2004; Lichtenstein, Fischhoff, & Phillips, 1982). In a typical laboratory task, participants are asked to report upper and lower values such that they are *X*% confident that the true value lies within this interval. *X* usually refers to fairly high levels of probability, such as 90%, 95%, or 98% (Bazerman, 1994; Klayman, Soll, Gonzáles-Vallejo, & Barlas, 1999). Similar procedures are adopted by textbooks in project management, where managers are encouraged to predict the most likely effort of a new project, along with a minimum–maximum prediction interval corresponding to $p = 0.9$ or higher (Kerzner, 2001; Moder, Phillips, & Davis, 1995). A common finding both in laboratory and real life is that people are poorly calibrated, as their hit rates rarely reach the assigned level of confidence. For 90% confidence intervals, hit rates around 50% appear to be the rule rather than the exception (Alpert & Raiffa, 1982; Connolly & Dean, 1997; Jørgensen, Teigen, & Moløkken, 2004; Juslin, Wennerholm, & Olsson, 1999; Klayman et al., 1999).

Most previous research on uncertainty intervals has been devoted to range judgments, where participants are asked to give lower and upper bounds that include the true value with *X*% confidence, for instance: “I am 80% sure that Charles Dickens was born between . . . . . . and . . . .” Soll and Klayman (2004) compared this procedure with a two-point format, where separate questions are asked about the higher and lower interval bounds, for instance: “I am 90% sure that Charles Dickens was born after . . . . . .,” and “I am 90% sure that Charles Dickens was born before . . . .” Although the range and the two-point format are formally similar (both are supposed to delimit an 80% confidence interval), Soll and Klayman found wider intervals, higher hit rates, and better calibration with the two-point method than with the range method.

Both these methods may be contrasted to the single-bound method, where only lower bounds or upper bounds are asked for and evaluated. Bolger and Harvey (1995) studied directional probability forecasts, where participants watched a trend and were asked to estimate the probability of data being above or below a specified reference value. Thus participants did not produce their own upper or lower bounds, but evaluated the likelihood of criterion values assigned to them by the experimenter. In this study participants appeared to be underconfident in their predictions.

Teigen and Jørgensen (2005) compared the range method and the single-bound method by asking participants to estimate the populations of two random samples of European capital cities. After the first set of 10 capitals, they were given feedback on their own performance, before receiving the second set of capitals. Participants in the range condition were asked to produce range estimates for both sets by completing sentences like: “I believe that London has between . . . . . . and . . . . inhabitants.” In two single-bound conditions, corresponding sentences were: “I believe that London has more than . . . . . .,” or “I believe that London has less than . . . . . . inhabitants.” It turned out that participants in the single-bound conditions

DOI: 10.1002/bdm

generated much wider intervals than participants in the range conditions (56% wider for the first set and 31% wider for the second set of capitals). They were also generally underconfident, whereas participants in the range conditions were overconfident even after feedback on their first set of responses.

We have not been able to find other studies in the research literature where the single-bound method has been systematically explored. This is surprising, as in daily life, single-point boundary estimates appear to be common. I promise to finish a piece of work "before the end of October," or to answer a letter "within 3 days." In these cases, information about upper bounds appears to be crucial, whereas the lower bounds may pass unmentioned. On other occasions, it is more natural to focus on lower bounds, as when I am warned that I have to pay "at least $10 000" for a bathroom renovation, or am told that the tsunami has claimed "more than 200 000 lives." Such estimates may be regarded as one-sided delimiters of an implicit uncertainty interval, or *one-sided uncertainty intervals*, for short.

Exclusive or inclusive interval bounds

When people are asked to produce uncertainty intervals, they can choose between two types of boundary terms, denoting exclusion or inclusion, respectively.

Exclusive values delimit an open interval, such that $X_1 < T < X_2$. The upper and lower values, X_1 and X_2 fall in this case outside the interval. For instance I can say that London has *more than* 1000 inhabitants or *less than* 50 millions without implying that these numbers are in any way realistic estimates of the population of London. Other exclusive terms are *above* versus *below*, and *after (later than)* versus *before (earlier than)*. For instance a person might say that Charles Dickens was born after 1800, or before 1850, without implying that Dickens was born in any of these years.

Boundary values can also be inclusive, delimiting a closed interval, such that $X_1 \leq T \leq X_2$. In this case X_1 and X_2 can be regarded as realistic, potential values. Terms denoting inclusive bounds are *at least/at most*; *maximum/minimum*; *as early as/as late as*; or *at worst/at best*. Results from a separate study (Teigen, Halberg & Fostervold, in press) showed that people clearly distinguished and made different interpretations of exclusive and inclusive terms. If someone says that a pair of shoes costs "minimum" or "at least" NOK 500 ($80), she implies that the shoes *could* in fact cost NOK 500. In contrast, "more than NOK 500" means that 500 falls outside of the price range. Inclusive estimates were, in this study, not only described as possible, but also as *probable* values, whereas exclusive boundary estimates were evaluated to be *improbable*.

The distinction between these two types of terms seems never to have been regarded as an issue in studies of uncertainty intervals. Interval bounds are in some contexts described as *min/max* values, or *worst case/best case* scenarios (inclusive terms), at other times they are described as values that the true value is supposed to lie *above* or *below*, or to be *more than/less than* (exclusive terms), with no attempt being made to investigate potential differences between these two classes of terms.

Exclusive intervals should in principle be somewhat wider than inclusive intervals, but the difference could be immaterial for most practical purposes. For instance, if a car travels with a speed of 30–40 mph, 30.0 falls inside and 29.9 falls outside of the range. I am in this case entitled to say that it had a minimum speed of 30 mph, or alternatively, that it went faster than 29 mph (or 29.9). This is an inconsequential difference, and seems to indicate that it does not really matter which term is being used. But the difference between being included and being excluded is psychologically important, and could lead to larger differences than the speed example suggests.

Yaniv and Schul (1997) asked one group of respondents to mark alternatives "that are likely to be the correct answers" (inclusion instructions), whereas another group should mark those "that are *not* likely to be the correct answers" (exclusion, or elimination instructions). The first group marked 18% "likely" alternatives. The second group marked 49.9% alternatives as "not likely," implying that the remaining 50.1% were likely. This difference suggests that people apply different criteria for inclusion and exclusion judgments (see also Yaniv, Schul, Raphaelli-Hirsch, & Maoz, 2002). People also seem to regard inclusive estimates as more informative and more precise than exclusive estimates (Teigen et al., in press).

DOI: 10.1002/bdm

Sources of uncertainty

Uncertainty intervals can be given two distinct interpretations, depending upon whether they describe uncertainty relating to a single, specific outcome or to a distribution of outcomes. The price of a particular product, or the number of seats in a particular aircraft, has one, and only one correct answer. When uncertainty intervals are suggested, they reflect an individual's degree of knowledge, with wide ranges suggesting inexact or imperfect knowledge. Alternatively, the task could be to describe the upper or lower limit of a particular class of products, like the price range for air tickets, or the number of passengers. Such ranges describe the distribution of values within each class. Wide ranges would, in these cases, suggest a high degree of outcome variability. Risk researchers have accordingly drawn a distinction between epistemic uncertainty and variability. Whereas epistemic uncertainty can be reduced through further research, variability is externally given and will not disappear with increasing knowledge (Morgan & Henrion, 1990).

This distinction is reminiscent of, but not identical to the traditional distinction between internal and external uncertainty. In their seminal paper on ''variants of uncertainty,'' Kahneman and Tversky (1982) drew attention to the fact that some probabilities must be taken to reflect an individual's state of knowledge (internal uncertainty), whereas others are attributed to chance factors and propensities inherent in the outside world (external uncertainty). My confidence of having submitted the correct answer in a multiple-question test belongs to the first of these categories, whereas the probability of drawing an ace from a well-shuffled deck of cards depends upon random events beyond my control, and can accordingly be described as reflecting external uncertainty. In theories of probability, these two concepts have been discussed as epistemic and aleatory probability, respectively (Hacking, 1975).

It is evident from these examples that uncertainty about single outcomes can be both internal and external. An individual's uncertainty about the current composition of the Norwegian parliament is clearly internal, and solely due his or her state of ignorance. Uncertainty about the composition of the parliament after the upcoming election can be due to both internal and external factors, by reflecting unpredictable processes in the society that cannot be controlled or foreseen. It will be shown in Experiment 2 that the same pattern of uncertainty intervals around single outcomes can be produced in both cases.

Uncertainty reflecting variability is typically more external than internal. For instance, my uncertainty about how much I have to pay for a flight to New York is only partly due to my general ignorance of prices, but even more to the fact that the prices of air tickets actually vary from very cheap to extremely expensive. A person with extensive knowledge of the distribution of prices would under these circumstances generate a wide uncertainty interval, perhaps even wider than a less knowledgeable person.

Most intervals that have been studied to date have been of the first kind. For instance, people are asked about specific historical or geographical facts (like the birth year of Charles Dickens, or the population of London). These questions are assumed to have one correct answer, so when an uncertainty interval is generated, it cannot reflect external fluctuations, but must be due to a lack of accurate information. Ideally, such uncertainty intervals should be as narrow as possible.

With uncertainty intervals due to a distribution of values, boundary values acquire a different meaning. Low minimum and high maximum values can in these cases reflect knowledge rather than ignorance. In domains characterized by variability, minimum and maximum estimates are accordingly more informative than in domains where there is only one correct answer, and all uncertainty must be attributed to incomplete information.

The present studies

In this article, we first present a study designed to compare price estimates generated by different elicitation methods: the *single-bound* method (where participants suggest either a lower or an upper value, but not both), the *two-bound* or two-point method (where the same participants suggest lower and upper values, in response to two separate questions), and the *range* method (where participants are asked to indicate an interval). Within the first two formats, we also compare the effects of inclusive versus exclusive terms, the question

DOI: 10.1002/bdm

being whether the differences between the different interval elicitation methods are primarily due to format or to choice of terms.

In the second study, we compare inclusive versus exclusive single-bound values to replicate the results from Experiment 1 in a different domain (composition of the Norwegian parliament before and after the elections).

In both these studies, wide uncertainty intervals reflect lack of precise knowledge of individual items. In Experiments 3 and 4 we compare inclusive versus exclusive intervals in areas where the uncertainty can be given a more concrete interpretation as being due to external variability. These studies disclosed, to our surprise, a reversal of the findings from the first two experiments.

Generation of upper and lower bounds may or may not correspond to people's interpretation of these bounds. In Experiment 5 participants are placed in the role of listeners, giving their best guess of what is implied by another speaker's lower and upper bound statements. What do they think is the most likely estimate when a person says that a specific hotel has *minimum* 100 rooms? Will people perceive a difference between inclusive and exclusive terms, or will they think ''*minimum* 100 rooms'' and ''*more than* 100 rooms'' can be used interchangeably, indicating the same quantity?

EXPERIMENT 1

Separate estimates of lower and higher interval bounds have been shown to produce wider intervals than corresponding range estimates. Such estimates can be given by the same subjects, who are asked two separate questions about higher and lower bounds (Soll & Klayman, 2004), or by different subjects, who are asked to produce either the higher bound, or the lower bound, but not both (Teigen & Jørgensen, 2005). The two-bound method, used by Soll and Klayman, and the single-bound method, introduced by Teigen and Jørgensen, have not been previously compared.

Several predictions are possible. Both methods ask for separate estimate of lower and higher bounds, directing the respondent's attention toward extreme cases. With the single-bound method, this effect could be reinforced, since attention is here exclusively focused on either high or low values. By this line of reasoning one would expect wider intervals with the single bound than with the two-bound method. On the other hand, Yaniv and Foster's (1995) informativeness hypothesis could lead to a different prediction. These authors have argued that people try to make their estimates as informative as possible. It can be argued that two-bound estimates are more informative than single bounds. A speaker, who predicts that a pair of shoes costs at least \$100, and at most \$200, has at the same time suggested a price around \$150 as a good guess. A single lower bound estimate (''at least \$100'') is less specific about the most likely price. To make such an estimate more informative, it must be pulled closer to one's best guess, leading to more narrow intervals.

Lower and upper bounds can be produced in response to inclusive as well as exclusive terms. In Soll and Klayman's (2004) study, both types of terms appear to be used, assuming that terms like ''at least'' and ''at most'' are inclusive terms, whereas ''after'' and ''before'' are exclusive. However, these terms were not compared, as different terms were used with different scenarios. In Teigen and Jørgensen's (2005) study, only exclusive terms (more than/less than) were used in the single-bound conditions. This makes it difficult to decide whether the wide uncertainty intervals found in this study were due primarily to the single-bound method or can be attributed to the use of exclusive terms.

In the present experiment these issues are investigated by asking separate groups to produce intervals based on the single-bound, the two-point, and the range method. Single-bound and two-point estimates are studied with both exclusive terms (more than and less than) and inclusive terms (minimum and maximum values), making direct comparisons possible. If inclusive terms are interpreted to mean low or high values that can be regarded as plausible, or even probable, we would expect narrow intervals (including only plausible

DOI: 10.1002/bdm

values). Exclusive terms might, in contrasts, initiate a search for values that are too low or too high to be considered plausible. This could give rise to intervals that are considerably wider.

In contrast with most previous interval studies, participants were not assigned a predefined level of confidence. Instead of assigned 90% or 99% intervals, participants were simply asked what they thought would be appropriate interval bounds. This method has been shown to lead to intervals of the same magnitude as assigned high-confidence intervals (Teigen & Jørgensen, 2005; Yaniv & Foster, 1997). Confidence estimates were subsequently collected in two ways: (1) Confidence ratings of each individual estimate, yielding ''local confidence'' (Liberman, 2004). (2) A final question about expected number of correct answers, which can be converted into a measure of ''global confidence.'' Previous research has shown that global confidence, based on estimated hits after all answers have been submitted, is often lower, and reveals less overconfidence than local confidence estimates (Gigerenzer, Hoffrage, & Kleinbölting, 1991; Griffin & Tversky, 1992; Liberman, 2004; Sieck & Arkes, 2005; Sniezek & Buckley, 1991; Teigen & Jørgensen, 2005).

Calibration studies have occasionally been criticized for using misleading, difficult, artificial, or biased samples of questions (Griffin & Brenner, 2004). In the present study, we ask students to estimate prices of items from a sports equipment catalogue. Price estimation was believed to be a relatively familiar activity for our participants, as most people in a western society have extensive shopping experience, which includes guessing, checking, evaluating, and comparing product prices. Sports equipment was chosen as a domain that would not selectively favor one sex, and was believed to include a range of products not too far removed from students' daily life, without being directly within their field of expertise. To avoid biased sampling, a representative set of items from a standard catalogue was included. This method is not equivalent to the ''natural sampling'' procedure advocated by Gigerenzer and Hoffrage (1999), where people are asked to give confidence estimates after having observed successive random draws from a pool of ''naturally'' defined items. Still, price range questions were assumed to be more natural for our participants than being exposed to haphazard collections of ''almanac'' questions often used in overconfidence research.

Method

Participants

Participants were 339 students following a course in introductory psychology at the University of Oslo (74% women, median age 21 years). They were randomly allocated to seven different conditions (A–G) with 40–50 participants in each group.

Questionnaires

All questionnaires contained short descriptions of ten items from a catalogue of sports equipment issued by a major Norwegian sports warehouse. Care was taken to select representative items covering the full price range of products, from cheaper items (basketball, life jacket) to more expensive ones (alpine skis, tent for three persons). Catalogue prices ranged from NOK 199 ($32) to NOK 2999 ($480), with a grand mean of NOK 1163 ($186). The prices were found to be comparable to prices for similar items from other dealers.

The questionnaires were introduced with some general information about the dealer (excerpted from the catalogue); participants were further informed that we had selected fairly typical products that were neither among the cheapest nor among the most expensive of their kind. Each item was accompanied by a 2–3 lines description from the product catalogue, for instance:

> Rucksack: *Salomon, HighRoute 500*. 50 litre backpack from Salomon for longer trips and mountain hiking. Equally well adapted for summer and winter usage. With a V-shape design and adjustable hip belt, it offers comfort as well as freedom of movement.

DOI: 10.1002/bdm

Interval estimates

(a) Single-bound conditions: Participants belonging to the four single-point estimate conditions were asked one of the following questions:

A: (low, inclusive): I think that the rucksack costs *minimum* (at least)[1] NOK ....
B: (high, inclusive): I think that the rucksack costs *maximum* (at most) NOK ....
C: (low, exclusive): I think that the rucksack costs *more than* NOK ....
D: (high, exclusive): I think that the rucksack costs *less than* NOK ....

(b) Two-point conditions: Participants in these conditions answered pairs of questions:

E: (inclusive): Questions A and B (as above).
F: (exclusive): Questions C and D (as above).

(c) Range condition:

G: I think that the rucksack costs *between* NOK .... and NOK ....

Confidence estimates. Participants in all conditions rated their confidence of each interval estimate by circling a number on an 11-point 0–100% certainty scale, one per item. These "local" confidence values were subsequently averaged over all items, yielding a mean local confidence score for each participant.

After completing the 10-item questionnaire they were asked to predict their total number of correct estimates ("I think I have .... correct answers"), indicating their global confidence in their performance.

Results

Some participants produced extreme high or low estimates, by answering for instance "between NOK 10 and NOK 100 000," or "less than one million." To prevent extreme cases to have a disproportionate influence on the averages, two procedures were adopted and compared. (1) Trimmed arithmetic means were calculated eliminating the largest 5% and smallest 5% of the cases (the two highest and the two lowest scores in each condition). (2) Extreme estimates above NOK 10 000 were arbitrarily set equal to 10 000. By this procedure, no observation was discarded, but the distribution was truncated by bringing 3.2% of upper bound estimates down within a more normal range (we saw no need for editing low values because of floor effects: extreme low values cannot be lower than zero).

Mean intervals for all items in all conditions are displayed in Table 1. For single-point conditions, intervals are based on the differences between lower trimmed means (in conditions A and C) and upper trimmed means (in conditions B and D). For two-point and range conditions, lower and upper trimmed means are based on estimates from the same subjects.

Two-point estimates led to intervals that were about 50% wider than estimates generated by the range method, confirming the findings of Soll and Klayman (2004). This difference cannot be attributed to choice of terms, as inclusive terms (Condition E) and exclusive terms (Condition F) yield similar intervals. Both are consistently wider than range estimates, for all ten items on the list.

Perhaps the most striking feature of Table 1 is the large difference between the two pairs of single-bound conditions. Exclusive terms (more than/less than) led to intervals that were twice as wide as inclusive terms (at least/at most). This massive difference is repeated for all ten items on the list and thus highly significant ($p < .002$ by a binomial test). For instance, snowshoes were believed to cost, on the average, *more than* NOK 370 and *less than* NOK 1611—a difference of NOK 1241. At the same time they were believed to cost *at least* NOK 570 (minimum price) and *at most* NOK 929 (maximum price)—yielding an interval of only NOK 359.

DOI: 10.1002/bdm

Table 1. Mean estimated price intervals (in NOK) for ten items of sports equipment, based on different elicitation methods

	Single-bound estimates		Two-point estimates		
	Inclusive B–A	Exclusive D–C	Inclusive E	Exclusive F	Range estimates G
Backpack	389	1030	776	983	477
Life jacket	346	892	502	490	313
Alpine skis	1640	2524	1834	1640	1522
Basketball	236	523	370	438	306
Rainwear	651	1296	766	692	461
Three-person tent	1362	2312	2462	1520	1254
Sleeping bag	566	1242	904	982	576
Snowshoes	359	1241	680	835	653
Child's bike seat	294	1257	722	689	416
Hiking trousers	449	1118	749	937	477
Mean intervals of:					
(1) trimmed means	629	1344	981	920	645
(2) reduced outliers	834	1965	1129	1093	849
Mean relative intervals computed from:					
(1) trimmed means	0.53	1.04	0.84	0.85	0.58
(2) reduced outliers	0.67	1.16	0.90	0.91	0.70

Note: All means for individual items are 5% trimmed means (where the largest 5% and the smallest 5% of the values within each condition have been eliminated).

The lower half of the table displays grand means for intervals calculated by both procedures: (1) trimmed means and (2) truncated distributions, where all observations are included, but a ceiling of 10 000 is introduced for extreme estimates. The second method led to wider intervals (as outliers are included), but the same general pattern was reproduced. Table 1 also shows mean relative intervals (MRI) based on both methods. Mean relative interval is defined as the ratio between mean interval width and the interval midpoint (Teigen & Jørgensen, 2005). Thus, the mean relative interval of a 200–400 range = (400–200)/300 = 0.67, to take one example. Relative intervals have the advantage over simple intervals by being standardized, making intervals of cheap and expensive items comparable.

Pairwise comparisons of relative intervals based on trimmed means for the ten individual items revealed highly significant differences between inclusive and exclusive single-bound estimates, $t(9) = 7.83$, $p < 0.001$, and between the two-point and the range conditions, $t(9) = 7.46$, $p < 0.001$ and $t(9) = 6.89$, $p < 0.001$, for inclusive and exclusive terms, respectively. The two-point method yielded significantly smaller intervals than the single-bound method for exclusive terms, $t(9) = 6.79$, $p < 0.001$, and significantly wider intervals for inclusive terms, $t(9) = 4.38$, $p < 0.002$. (Similar analyses of relative intervals based on truncated distributions, and on mean intervals computed by either method, gave also significant results for all these differences.) However, inclusive and exclusive estimates generated different intervals only in the single-bound conditions. With the two-point method, no difference between the two types of terms was observed. A schematic comparison of the boundary values generated by different elicitation methods is presented in Figure 1.

Hit rates were generally higher in the single-bound conditions than in the two-point and range conditions. This was to be expected, as single-bound estimates can miss the correct value in only one way (the correct value can be below the lower limit, or above the higher limit, depending on condition), whereas two-point and range estimates are more risky in the sense that two types of misses are possible in the same condition. As shown in Figure 2, hit rates were around 70% in the single-bound conditions, close to 50% in the two-point

DOI: 10.1002/bdm

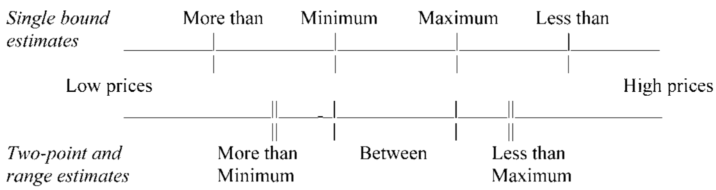

Figure 1. Estimated price ranges suggested by different elicitation methods, Experiment 1

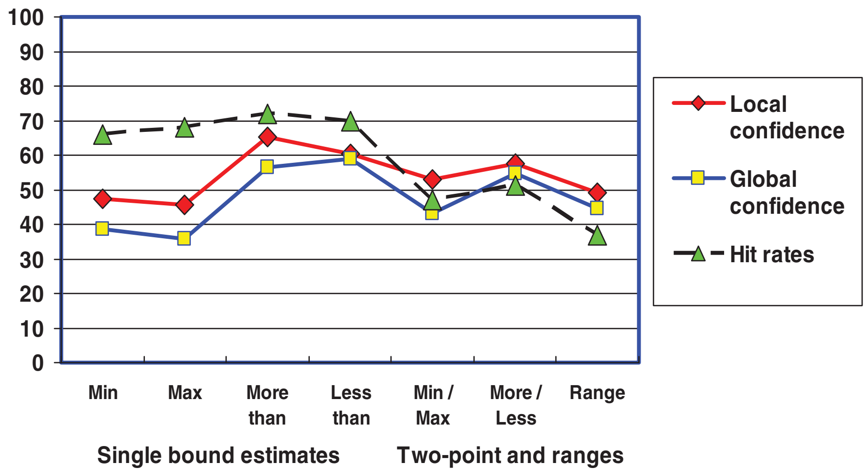

Figure 2. Confidence estimates and hit rates in seven conditions, Experiment 1

conditions, and down to 37% in the range condition, where estimated intervals were narrow and easily missed.

A comparison of hit rates and confidence estimates shows that participants were underconfident in the single-bound conditions and overconfident in the range condition (Figure 2). Local confidence was, as predicted, somewhat higher than the global confidence (grand means of all conditions = 53.7% for local confidence and 46.9% for global confidence, $t(324) = 6.13$, $p < 0.001$).

Confidence estimates differed between the conditions. Figure 2 shows that exclusive estimates consistently implied higher confidence than inclusive estimates. Mean local confidence is 61.2% for exclusive terms, averaged over three conditions, against 48.6% for inclusive terms in a parallel set of conditions. A 2×3 ANOVA showed a highly significant main effect of term (inclusive vs. exclusive), $F(1, 269) = 24.2$, $p < 0.001$, no effect of type of estimate (lower bound, higher bound, and two-point), and a significant interaction $F(2, 269) = 3.11$, $p < 0.05$, reflecting the fact that the difference between exclusive and inclusive estimates appears to be smaller in the two-point condition. Global confidence estimates reveal an even larger effect of term, as participants in the three exclusive conditions thought they had, on average, 5.67 correct responses against 3.92 in the three inclusive conditions. A two-way ANOVA for these conditions shows a highly significant main effect of term, $F(1, 277) = 32.1$, $p < 0.001$, no effect of type of estimate, and no significant interaction. Thus people seem more comfortable about their ‘‘more than’’ and ‘‘less than’’ price estimates, than they do about their ‘‘minimum’’ and ‘‘maximum’’ prices. This is confirmed by the higher hit rates in the exclusive conditions.

DOI: 10.1002/bdm

Discussion

Experiment 1 shows that the boundaries of an uncertainty interval can be strongly influenced by elicitation format. We replicated Soll and Klayman's (2004) finding of wider intervals by the two-point method than the range method, and Teigen and Jørgensen's (2005) wide intervals when exclusive terms are used as single bounds. A comparison of the two-point and single-bound method revealed, somewhat to our surprise, that both predictions came true: Single-bound intervals can be both wider and tighter than two-bound intervals. It all depends upon whether the bounds are based on exclusive or inclusive terms. When people are asked to provide minimum or maximum values, they seem to make an attempt to be as informative as possible, leading to narrow intervals. When they are asked to suggest exclusive values, which are either lower or higher than *T*, they go to the other extreme and provide low and high values that are very far apart. This agrees well with results from a separate study (Teigen et al., in press) which indicated that inclusive and exclusive interval bounds can be interpreted as likely versus unlikely values.

The results also replicated Teigen and Jørgensen's (2005) finding that people are not always "overconfident" about their uncertainty intervals. Overconfidence is typically found when people are asked to produce intervals corresponding to a specific (usually high) level of confidence. When the confidence is evaluated, rather than assigned, by probability estimates corresponding to a specific interval, it can be brought down to a more realistic level (see also Winman, Hansson, & Juslin, 2004), and even reversed. Participants in the present experiment were overconfident in the range condition, where intervals were narrow, but underconfident in their single-bound estimates. It is perhaps more correct to say that confidence estimates varied less than the hit rates. This is to be expected if confidence estimates reflect the participants' assessment of their level of knowledge, more than their chances of hits. Chances of hits are clearly better with one-sided than with two-sided intervals, whereas the level of knowledge remains much the same.

EXPERIMENT 2

The large difference between inclusive and exclusive bounds in a context of price estimates called for a replication study in a different domain. In Experiment 2, students were asked to estimate the composition of the Norwegian parliament, 2 weeks before the parliamentary election in September 2005. The Norwegian parliament consisted at that time of representatives from eight different political parties, holding 1–43 seats of a total of 165. We assumed that most students had an idea of the relative sizes of these parties, without knowing the exact number of MPs belonging to each party. They were further asked to predict the new distribution of representatives after the upcoming election.

Method

Participants

Participants were 216 students following a course in introductory psychology at the University of Oslo (73.6% women, median age 21 years). They were randomly allocated to four different conditions of approximately the same size.

Material

The questionnaire consisted of two pages. The first page listed alphabetically the seven largest political parties in Norway, asking the participants to estimate their current number of seats in the Norwegian parliament. They were also informed about the total number of seats (165).

DOI: 10.1002/bdm

Estimates were given in conditions A–D in four different ways:

A: The Labor party has *minimum* . . . . representatives
The Progress party . . . .
The Conservatives . . . . . .
The Christian Democratic Party . . . .
The Centre Party . . . .
The Socialist Left Party . . . .
The Liberals . . . .
B: The Labor party has *maximum* . . . . representatives
. . . .
C: The Labor party has *more than* . . . . representatives
. . . .
D: The Labor party has *less than* . . . . representatives
. . . .

Following each estimate, participants were asked to indicate their confidence by checking a number of an 11-point 0–100% probability scale (Local confidence). Finally, they were asked about their estimated number of correct answers (Global confidence), and to rate how difficult they felt the task had been on a scale from 1 (very easy) to 7 (very difficult).

On the second page, the same parties were listed again, the task being to predict their number of seats after the election. This question was also stated in terms of *minimum*, *maximum*, *more than*, and *less than* values as above. Each statement was followed by confidence estimates, with a final question about how many of the seven statements that they believed they had answered correctly.

Results and discussion

Three participants (all in condition D) gave extreme estimates, by stating that all parties had less than 200, 1000, or 1 million seats in the Parliament. As the total number of seats is 165, these estimates were reduced to 165, to prevent outliers from having a disproportionate effect on the overall mean. Interval widths for inclusive and exclusive intervals were computed as in Experiment 1 by subtracting mean lower bound estimates from mean higher bound estimates, for each of the seven political parties. The results are presented in Table 2. The table shows that the exclusive intervals are about three times wider than the inclusive intervals. Uncertainty intervals for future election results were somewhat wider than uncertainty intervals for the present number of seats. A two-way repeated measures ANOVA for the values entered in Table 2 shows a

Table 2. Single-bound intervals based on mean estimated numbers of MPs from seven political parties, Experiment 2

	Before election		After election	
Parties	Inclusive B–A	Exclusive D–C	Inclusive B–A	Exclusive D–C
The Labor Party (Ap)	6.9	23.5	10.9	28.8
The Progressive Party (Frp)	5.4	22.5	8.6	25.5
The Conservative Party (H)	4.1	19.3	8.8	28.0
The Christian Democratic Party (Krf)	8.9	17.2	12.0	28.0
The Centre Party (Sp)	7.6	20.1	9.8	27.4
The Socialist Left Party (SV)	5.1	23.1	10.5	20.2
The Liberals (V)	9.4	23.0	10.1	20.1
Mean intervals	6.7	21.2	10.2	25.4

Note: Intervals in the inclusive condition are based on the difference between mean Minimum and Maximum estimates (conditions A and B); intervals in the exclusive condition are on differences between the mean More than–Less than responses (conditions C and D).

DOI: 10.1002/bdm

highly significant effect of elicitation method, $F(1,6) = 278.2$, $p < .001$, and a significant effect of time (present vs. future), $F(1,6) = 11.7$, $p < 0.05$, and no significant interaction.

A closer inspection of mean estimates shows that inclusive and exclusive lower bound estimates are of the same magnitude. This may be partly due to a floor effect, as some parties are quite small (the smallest occupying only two seats). In contrast, the upper bound estimates of conditions B and C are widely different from each other, *less than* estimates being much higher than *maximum* estimates. These differences are significant both for estimates of the present number of seats and for the predicted number after the election. Thus, the differences in interval width presented in Table 2 are chiefly due to different upper bound estimates.

Exclusive intervals led to higher hit rates, as predicted. Participants were also more confident in their *less than* estimates than in their *maximum* estimates, as in the previous experiment. Global confidence estimates were slightly lower than mean local estimates (the difference being significant only in the prediction conditions). Both these confidence estimates were considerably lower than the actual hit rates. Overall mean for local confidence estimates was 51.1%, the mean global confidence estimate was 47.3%, against an overall hit rate of 63.7%. Thus participants believed they had only 3.5 correct answers, but had generally 4.5 correct answers out of seven; they were in other words clearly *underconfident*.

Difficulty ratings revealed that high bound ratings were considered more difficult than low bound ratings (mean ratings 5.27 vs. 4.55). A 2×2 ANOVA revealed a significant main effect of lower vs. upper bounds, $F(1, 203) = 9.36$, $p < 0.01$, a non-significant main effect of inclusive vs. exclusive bound, and a significant interaction, $F(1, 203) = 4.45$, $p < 0.05$. The interaction is due to the fact that not all high bound estimates are equally difficult; *less than* estimates being considered easier than *maximum* estimates (mean difficulty ratings of 4.96 and 5.68 for less than and maximum estimates, respectively). Our interpretation is that *maximum* values require careful consideration since they are supposed to indicate the highest *likely* value, whereas *less than* values can be allowed more latitude as they are expected to fall outside of the range of likely values. This means that any value above the "maximum" is acceptable, requiring less deliberation. For instance, if I expect the Labor party to grow from 43 to about 50 seats, it is difficult to set an upper limit corresponding to their maximum, but much easier to decide that it will be less than an arbitrary high number, like 80 or 100.

EXPERIMENT 3

Both previous studies asked participants to produce an uncertainty interval around a single target value (the price of a particular product, and the number of seats belonging to a particular party). But intervals can also be used to describe a distribution of items. Minimum and maximum prices for stocks, concert tickets, and flights can be far apart because these prices are not fixed and can be high or low, dependent upon quality, date, and market factors like supply and demand. In the present experiment we asked participants to estimate upper and lower bound for the prices of flight tickets, a domain where prices are known for some degree of variability. This could lead even informed individuals to suggest wide intervals, delimited by low minimum and high maximum values. Yet, exclusive intervals should be still wider. If the maximum price for a flight to Rome is assumed to be $1000, it makes sense to say that it will cost "less than $1500." It makes less sense to say that it will cost "less than $500." However, to our surprise, the results told a different story.

Method

Participants

Participants were 89 students following an introductory course at the University of Oslo (82% women, median age 23 years). They were randomly allocated to two conditions, 1 ($n = 45$) and 2 ($n = 44$).

DOI: 10.1002/bdm

Material

All participants were given a questionnaire where they were asked to estimate approximate economy class round trip flight prices from Oslo to six different destinations. In Condition 1, they were asked to give exclusive lower bound prices for three destinations (Rome, Chicago, and Uzbekistan), by completing statements of the form: ''A flight from Oslo to Rome and back will cost more than NOK . . . .'' For three other destinations (Madrid, St. Petersburg, and Fiji) they were asked to give exclusive upper bound prices, for example, ''A flight from Oslo to Madrid and back will cost less than NOK . . . .'' The destinations in both sets were matched for price level; mean actual prices in Set I were NOK 3300 (Rome), 4600 (Chicago), and 20 000 (Uzbekistan), and in Set II: NOK 3100 (Madrid), 5100 (St Petersburg), and 20 000 (Fiji).

Participants in Condition 2 received the same six items, but were asked for inclusive bounds (minimum or maximum prices). Half of the participants in each condition gave lower bound estimates before higher bound estimates, for the other half, the order was reversed.

On the second page of the questionnaire they were asked to give the most likely price they would have to pay for a round trip to these six destinations.

Results

Table 3 reveals a pattern of estimates that is very different from the previous two experiments. The table is based on answers from 88 participants; one outlier was removed (this participant answered ''more than 0,'' or ''less than 50 000,'' to all questions). Lower limit estimates were quite similar in the two conditions (no significant differences), whereas upper limit estimates led to much higher values in the inclusive (maximum) condition than in the exclusive (less than) condition. (These results remain significant even if the outlier is included.) From this we can infer that in the domain of flight prices, estimates of inclusive intervals will be *wider* than estimates of exclusive intervals.

This is confirmed from the results of the second page of the questionnaires, where participants were asked to give the most likely mean prices. Distances between estimated means and estimated lower limits (page 2 vs. page 1 prices) were larger for inclusive estimates than for exclusive estimates in two out of three cases (Rome and Chicago), and the distances between means and upper limits were larger for inclusive estimates in all three cases.

This finding makes sense when we realize that flight prices actually show a good deal of variability, depending upon airline, time of week, budget offers, etc. Thus low minimum and high maximum prices can in this case reflect external variability, whereas the minimum/maximum ranges produced by participants in Experiments 1 and 2 reflect their general ignorance of the ''true'' numbers. The price of a particular backpack, and the number of MPs from a specific political party are not the subject of fluctuations. Thus the minimum and maximum estimates should stay as close to a single, most likely value as possible. With a wide

Table 3. Mean inclusive and exclusive interval bounds for flight prices (in NOK), Experiment 3

	Inclusive	Exclusive	t	p
Lower bounds	Minimum	More than		
Oslo–Rome	1662	2049	1.32	ns
Oslo–Chicago	4291	4644	0.44	ns
Oslo–Uzbekistan	5576	5074	0.38	ns
Higher bounds	Maximum	Less than		
Oslo–Madrid	6962	3386	5.74	.0001
Oslo–St. Petersburg	8280	4084	5.48	.0001
Oslo–Fiji	15 390	9530	3.83	.001

DOI: 10.1002/bdm

distribution of target objects, questions about minimum and maximum values might, in contrast, serve as reminders about the possibility of very small and very high values.

EXPERIMENT 4

Experiment 3 showed that inclusive terms do not always lead to narrow intervals. Minimum and maximum prices were in this study further apart than mean estimates of *more than* and *less than* values. This was a surprising finding, as inclusive intervals by definition should be surrounded by exclusive values, rather than the other way around. It also runs counter to the results of Experiments 1 and 2, where minimum and maximum values were less extreme. In these experiments, participants estimated specific, singular items, whereas Experiment 3 asked for flight prices, which are known for being variable.

To test the notion of variability, we designed an experiment which asked for single-bound estimates of variable as well as non-variable items within the same general domains. For instance a question about the number of rooms in one specific hotel has only one correct answer. In this case we would expect estimates of the *minimum* and the *maximum* number of rooms to come close together, creating a narrow (perhaps too narrow) interval. We would further expect exclusive *more than* and *less than* estimates to be more extreme, creating wider intervals, in line with the results from Experiments 1 and 2.

If we ask instead how many guests will be staying at this hotel at some unspecified point of time during the winter season, participants will realize that this is actually a fluctuating number. The minimum number of guests could be quite low and the maximum number could be quite high, depending upon random factors, day of week, as well as conferences and other events attracting visitors or keeping them away. Questions about minimum and maximum numbers might in this case serve as a kind of priming procedure, prompting the participants to think about very quiet versus very busy days, respectively. Exclusive ''more than'' or ''less than'' estimates might be less affected by day-by-day fluctuations, as they do not in the same way indicate the number of actual guests, but refer to external numbers lower or higher than the most likely target values.

Estimates of stable values, like the number of rooms, should accordingly lead to a replication of the results found in Experiments 1 and 2, whereas estimates of *fluctuating* quantities, like the number of guests, should lead to wider intervals, especially for inclusive boundary terms. This could neutralize or perhaps even reverse the difference between exclusive and inclusive boundary terms that was demonstrated in the first two experiments.

Method

Participants

Participants were 334 students at the University of Oslo (68% females, median age 22 years). Of these, 142 were students following a course in introductory psychology (students belonging to this course were randomly assigned either to Experiment 2 or to the present experiment). The remaining 192 participants were recruited on campus and asked to answer a questionnaire containing the questions below, together with several other, unrelated judgment tasks, and were paid with lottery tickets for their participation. Participants in both samples were randomly allocated to eight sub-conditions, with about 40 in each condition. The general trends in both samples were the same, so the data were pooled.

Material

Two sets of items were prepared. Set I required participants to estimate interval bounds for *specific* target objects, which were assumed to be of a definite magnitude. Set II items described *variable* target objects (distributions of targets), belonging to the same domains.

DOI: 10.1002/bdm

Set I: Specific targets:

- Number of tables at the Theatre Café [traditional downtown restaurant in Oslo]
- Number of seats in the streetcar running between downtown and the State Hospital
- Number of rooms (single and double) in Grand Hotel [most famous hotel in Oslo]
- Average monthly rent for university students
- Average weekly allowance for Norwegian 14-year olds
- Average student debts for law students in Oslo at the time of their graduation

Set II. Variable targets:

- Number of patrons visiting the Theatre Café during a weekday evening
- Number of passengers boarding the streetcar between downtown and the State hospital around noon
- Number of house guests in Grand Hotel, a night during the winter season
- Monthly rents for university students
- Weekly allowances for Norwegian 14-year olds
- Student debts for law students in Oslo at the time of their graduation

It was assumed that Set I target objects would be seen as having only one true answer. For instance, there are 207 rooms in Grand Hotel, so when a participant suggests that there must be more than 100 rooms, or less than 300, the interval must be attributed to the individual's lack of precise knowledge, and not to variability in the actual number of rooms. Similarly, the average rent for a student flat refers to a specific mean value, based on a distribution of individual rents. In contrast, the number of house guests at Grand Hotel will vary from night to night, and the rent paid by individual students will vary from cheap to quite expensive, creating a true *range* of guests and prices, respectively. A low minimum, or a high maximum value, could in this case reflect beliefs about variability rather than lack of knowledge.

Within both conditions, four separate groups were asked to estimate *minimum* values, *maximum* values, *more than* values, and *less than* values, creating a $2 \times 2 \times 2$ between-subjects design, the factors being fixed versus variable target objects (Set I vs. Set II), inclusive versus exclusive boundary terms, and lower versus upper bounds. All answers were of the form "I think there is more than [less than; minimum; maximum]... tables [café guests]". There were no instructions about the exact level of confidence that should be attained. Previous research (Teigen & Jørgensen, 2005) has shown that different verbal phrases (like "I believe that ...," or "I am certain that ...") will lead to similar uncertainty intervals, even if they seem to imply different levels of confidence.

Results and Discussion

Mean estimates of Set I values replicated the response pattern from Experiments 1 and 2. Table 4 shows that all inclusive upper values lie below the exclusive upper values, and all but one of the inclusive lower values lie above the exclusive lower values, creating more than/less than intervals that are in all cases wider than the minimum/maximum intervals ($p < .05$, sign test). Table 4 displays trimmed means where the 5% highest and 5% lowest estimates have been discarded, to prevent outliers to have a disproportionate influence on the results. This general pattern can be reproduced from untrimmed means as well as from median bounds. Intervals are more similar for estimated means (the last three items) than for estimated number of tables, seats, and rooms. Apparently some participants did not clearly distinguish between estimating an uncertainty interval for a mean and estimating the range of individual observations upon which the mean is based.

Set II estimates reversed this pattern. The results reported in lower half of Table 4 show minimum/maximum intervals to be in all cases wider than the more than/less than intervals ($p < .05$, sign test). It appears that inclusive estimates (minimum–maximum values) convey a specific and concrete meaning in domains with external variability. Questions about minimum values may have activated thoughts about empty trams and children with low weekly allowances, whereas questions of maximum values make the

DOI: 10.1002/bdm

Table 4. Mean single-bound estimates of fixed and variable quantities, Experiment 4

	Inclusive terms			*Exclusive terms*		
	Minimum	Maximum	Interval[a] (MRI)[b]	More than	Less than	Interval[a] (MRI)[b]
Set I (Fixed)						
Café tables	36	78	42 (0.73)	32	87	**55 (0.92)**
Streetcar seats	78	107	31 (0.34)	61	123	**62 (0.67)**
Hotel rooms	181	285	104 (0.45)	164	405	**341 (1.20)**
Mean rent (NOK)	2800	3669	889 (0.27)	2606	3984	**1378 (0.42)**
Mean allowance	135	318	183 (0.81)	123	361	**238 (0.98)**
Mean debt	263 827	360 393	96 566 (0.31)	266 752	370 370	**103 618 (0.33)**
Set II (Variable)						
Café patrons	101	229	**128 (0.78)**	277	286	9 (0.03)
Tram passengers	70	520	**450 (1.53)**	350	302	−48 (−0.15)
Hotel guests	71	206	**135 (0.97)**	138	211	73 (0.42)
Rents (NOK)	2155	4686	**2531 (0.74)**	2552	4551	1999 (0.56)
Allowances	74	315	**241 (1.24)**	129	343	214 (0.91)
Student debts	206 694	414 167	**207 473 (0.67)**	250 380	449 787	199 407 (0.57)

Note: All means for individual items are 5% trimmed means (where the largest 5% and the smallest 5% of the values within each condition have been eliminated).
[a]Intervals = Mean upper bound − mean lower bound. The larger intervals are printed in bold.
[b]MRI (Mean relative intervals) = Intervals/Interval midpoint.

notion of crowded trains and spoiled kids more accessible, perhaps functioning as a kind of anchor for the subsequent estimate. ‘‘More than’’ or ‘‘less than’’ questions, being exclusive, do not in the same way indicate low and high target objects, and do not prime a high degree of variability. If anything, the concept of ‘‘more than’’ may suggest objects of a higher value than the concept of ‘‘less than,’’ even if these concepts in themselves are assumed to indicate lower versus upper limits, respectively. Otherwise it is difficult to explain how one lower limit mean (of tram passengers) could become *higher* than the corresponding upper limit mean. This rather anomalous (but non-significant) result can of course also be due to random group differences, as the two means were estimated by separate groups of participants.

To make the estimates from different tasks comparable, MRI were computed by dividing mean intervals with interval midpoints. MRI values for individual items are displayed in parentheses in Table 4. MRI for Set I, inclusive intervals, yielded a grand mean of 0.49, against 0.75 for Set I exclusive intervals. In Set II the reverse relationship holds, with a grand mean of 0.96 for inclusive values, against 0.35 for exclusive values. A 2×2 repeated measures ANOVA for the MRI values in Table 4 revealed no main effects, but a significant interaction between Set and boundary terms, $F(1,5) = 9.085$, $p < 0.05$, confirming the opposite pattern of Set I and Set II intervals. Grand means for the two sets of items are displayed in Figure 3.

EXPERIMENT 5

We have shown that people tend to create smaller uncertainty intervals with inclusive single-bound estimates than with exclusive single-bound estimates, when guessing at non-variable quantities, like the price of a particular object and the number of tables in a particular restaurant. With a distribution of variable quantities, like patrons visiting a restaurant, the opposite may occur. The question raised in the final experiment is whether these effects are predicted by listeners. Are people aware that minimum/maximum estimates sometimes imply tighter intervals than more than/less than estimates, and wider under other circumstances? If the results of the previous experiments reflect common knowledge, people should believe that the ‘‘true’’

DOI: 10.1002/bdm

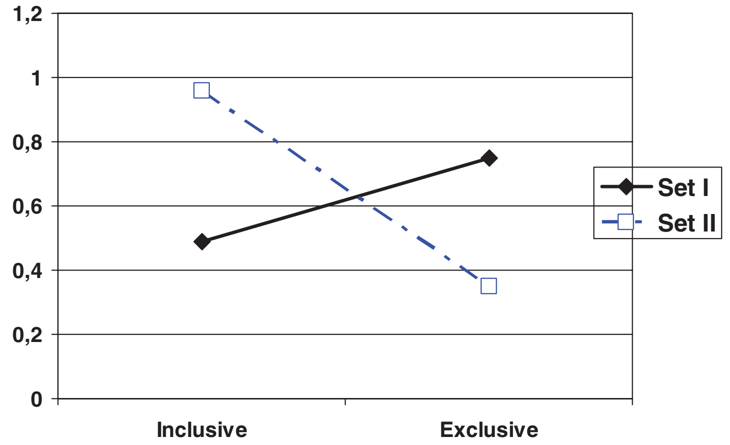

Figure 3. Mean relative intervals (MRI) based on inclusive and exclusive single bounds. Grand means for specific targets (Set I) and variable targets (Set II), Experiment 4

number of tables in a restaurant is not so far from the minimum/maximum estimate, but could be quite a bit different from estimated more than/less than values.

In the present experiment participants were asked to guess the most likely value a speaker has in mind when stating one-sided uncertainty intervals limited by inclusive or exclusive higher or lower bound phrases. The target objects to be described were identical to the first three items of Set I and Set II in Experiment 4.

Method

Participants

Participants were 268 students following an introductory course in psychology. Only students who had not already participated in Experiment 4 were included. They were randomly allocated to eight different conditions.

Material

Participants were given the first three items from Set I or from Set II, and asked to imagine (Set I) that someone estimated the number of tables in the Theatre Café to be ‘‘minimum 30;’’ ‘‘more than 30;’’ ‘‘maximum 80;’’ or ‘‘less than 80.’’ Each participant received only one question per item. The corresponding questions for Set II, given to separate groups of participants, would be to estimate the number of guests as ‘‘minimum 100;’’ ‘‘more than 100;’’ ‘‘maximum 200;’’ or ‘‘less than 200.’’ The numbers were in all cases chosen to be close to the median responses of Experiment 4. Participants were asked in each case to state which number they thought the speaker had in mind as his/her ‘‘best guess.’’

Design

The study paralleled Experiment 4 by being a $2 \times 2 \times 2$ between-Ss experiment, the factors being fixed versus variable target objects (Set I vs. Set II), inclusive versus exclusive boundary terms, and lower versus upper bounds.

DOI: 10.1002/bdm

Table 5. Mean relative semi-intervals (MRSI) estimated from lower and upper bounds, Experiment 5

	Inclusive bounds		*Exclusive bounds*	
Based on:	Lower MRSI Minimum	Upper MRSI Maximum	Lower MRSI More than	Upper MRSI Less than
Set I				
Café tables	0.30	0.23	0.27	0.14
Streetcar seats	0.20	0.11	0.17	0.18
Hotel rooms	0.23	0.13	0.19	0.16
Mean Set I	0.24	0.16	0.21	0.16
Set II				
Café patrons	0.22	0.19	0.28	0.27
Tram passengers	0.16	0.35	0.15	0.44
Hotel guests	0.26	0.24	0.38	0.23
Mean Set II	0.21	0.26	0.27	0.31

Results

This study corresponds to Experiment 4 where people were asked to produce upper and lower bounds based on their ideas of the number of tables, guests etc. In Experiment 5 participants are instead given the upper and lower bounds and asked what these values imply. If they perceive exclusive and inclusive bounds to indicate intervals of different widths, this should be reflected in their estimates of the speakers' best guesses.

To make guesses of different target items comparable, we calculated mean relative semi-intervals (MRSI), defined as the difference between boundary value and best guess, relative to best guess. Thus if "minimum 30 tables" implies "40 tables" as one's best guess, MRSI = (40–30)/30 = 0.33. Lower and upper MRSI can be added to form an equivalent of MRI (see Experiment 4). Nineteen responses of a total of 794 (2.4%) were discarded because the best guess deviated from the boundary value in the wrong direction (as for instance if "more than 75 passengers" were said to imply a best guess of "50 passengers," or another number below the lower limit). Some of these mistakes could probably be attributed to cases where the participants disagreed with the stated limits. The MRSI values for the remaining estimates are shown in Table 5.

It is apparent from the table that participants in this experiment did not discriminate between inclusive and exclusive estimates. They believed that boundary limits would be placed on the average 20–25% away from the best guess regardless of the terms used. If anything, inclusive intervals were slightly wider than exclusive intervals for Set I items and slightly narrower for Set II items, going in the opposite direction of what would be predicted from Experiment 4. Thus listeners seem to be unaware of the effects of boundary terms on interval width, as demonstrated in the previous experiments. Best guesses seem to be produced simply by adding or subtracting a certain fraction from the estimated lower or upper bounds, perhaps by an anchoring and adjustment process, with little consideration as to how these estimates are produced. This could in practice lead to misunderstandings, for instance a price estimate of "more than $100" or "minimum $100" would be given the same interpretation, whereas Experiment 1 indicates that a product that is estimated to cost "more than $100" could at the same time be estimated to have a "minimum" price of $200 or more.

GENERAL DISCUSSION

The present studies were designed to study determinants for one-sided uncertainty intervals, where people are asked to estimate either lower bound values, or upper bound values, but not both. We have focused on two factors that have not been previously explored, namely (1) the nature of boundary terms: inclusive or

DOI: 10.1002/bdm

exclusive, and (2) estimation target: single objects or distributions. With single objects the source of uncertainty will be incomplete information, with distributions uncertainty is largely due to variability.

The two first experiments showed that minimum/maximum estimates of single objects led to tighter intervals than more than/less than statements. This was to be expected from a linguistic analysis of the terms involved, as inclusive (closed) intervals where the boundaries belong to the interval should, by definition, be smaller than exclusive (open) intervals where the boundary values fall outside of the interval. However, the exclusive intervals in the present studies were not simply larger, they were twice as large. This indicates that our participants were not simply trying to locate the dividing point between likely and unlikely values, with slight adjustments for inclusive versus exclusive values, but seemed to treat the questions of inclusive versus exclusive boundary values as very different tasks.

Experiments 3 and 4 showed, to our surprise, that this relationship can be reversed. When participants are asked to estimate minimum or maximum prices for flight tickets, or minimum and maximum number of hotel guests, they are seeking the extreme values of a distribution of target values. This led to very wide intervals. Questions about exclusive interval bounds—prices and numbers that the target value must be over, or under—did not in the same way encourage extreme estimates. Thus, estimated exclusive intervals became, paradoxically, tighter than the inclusive estimates. Again, type of terms seem to have change the nature of the task, but with opposite results.

The present studies were not designed to provide a theoretical explanation of these puzzling findings, and we can only offer speculations about their source and nature.

When people are asked to produce *inclusive* lower or upper bounds, for instance by indicating minimum or maximum values, we assume that they will start a search for possible exemplars that are as low as, or as high as these values. When very low or very high instances cannot be imagined, or retrieved from memory, the interval delimited by these boundary terms will become a tight one.

It is easy to see how such intervals can be influenced by the source of the uncertainty. When one is guessing at the minimum or maximum value of variable quantities, for example, October temperatures in London, it is possible to imagine quite cold days as well as very mild or warm ones, pushing minimum and maximum temperatures far apart. Questions about minimum temperatures may in themselves function as a kind of priming procedure, making cold days especially accessible in memory, whereas questions about maximum temperatures can activate memories and fantasies about late Indian summer days. This process can be conceived in analogy to anchoring procedures, which according to recent theories (Mussweiler & Strack, 1999) are assumed to work because low and high anchors selectively activate different evidence for the target object. When the question is to estimate single lower and upper interval bounds, one may not even bother to think about middle values, like the most likely, or normal temperature, the primary focus being on extreme instances.

When, in contrast, one is asked to guess the minimum or maximum value of a single, non-variable object, for instance the temperature yesterday at noon, it is more natural to start with a search for the most likely value, and to create an uncertainty margin around this value that will encompass potentially lower or potentially higher values. With inclusive terms, this interval will be limited by the fact that only *likely* values are included.

Most studies of uncertainty intervals seem to have used non-variable target values, like general knowledge questions (which, despite the term, turn out to be questions about specific facts), and time and effort estimates of specific tasks (Connolly & Dean, 1997; Jørgensen et al., 2004; Teigen & Jørgensen, 2005). This was also the case in the two first experiments of the present paper, where we asked for price estimates of specific items and for the number of MPs from specific political parties. In such cases, people have to imagine minimum or maximum values for one single object. This will naturally lead to a restricted interval. With perfect knowledge, the minimum and maximum should coincide with the most likely, or "true" value. Under conditions of relative ignorance, there will still be strong demands to place the minimum and maximum estimates as close to the true value as possible.

DOI: 10.1002/bdm

Exclusive values tell a different story. An exclusive low value is a value that falls below the plausible range, whereas an exclusive high value is a value that is too high to be true. This should encourage a search for unlikely or implausible values rather than likely ones. Such an approach could in principle lead to very extreme estimates. However, most participants apparently want to give relevant and informative estimates, as discussed by Yaniv and Foster (1997). Yet our first two experiments clearly showed that exclusive higher and lower bounds produced much wider uncertainty intervals than those that were delimited by inclusive terms.

Experiments 3 and 4 showed that for variable targets, inclusive estimates led to much wider intervals, in line with the analysis above. It is harder to explain why exclusive bounds should be less extreme. Perhaps "more than" and "less than" instructions are less effective primes than "minimum" and "maximum" instructions for activating thoughts about low and high instances, respectively. The reason may simply be that exclusive terms do not directly describe the lowest or highest imaginable target objects. They rather deny the existence of target objects corresponding to values as low as, or as high as the suggested bounds. One could even speculate that the phrase "more than" has a potential to prime high target objects, whereas "less than" can suggest low target objects, because of the connotations of the terms "more" and "less." Recent results (Teigen, 2006) indicate that people expect "more than" to describe an increasing, and "less than" a decreasing quantity. This could lead to occasional reversals of lower and upper bounds, as one of the items in Experiment 4 seemed to indicate.

The phrases used in this study do not exhaust people's lexicon of boundary terms. For instance, "not less than" and "not more than" are frequently used expressions of lower and upper limits. In prediction contexts "best case" and "worst case" scenarios are common. Advertisements often describe one tail of the price range for a line of products as "prices *from* . . . " (the upper limit being less appealing and goes unmentioned). All these expressions appear to be inclusive, in the sense that they represent possible values. Exclusive values may be described as prices "higher than" or "lower than" a certain value, a work will be finished "within" 10 days, or "earlier" or "later" than a certain date, and an outcome can be "better" or "worse" than expected.

Single-bound phrases do more than providing quantity information. They allow speakers to describe the same event from different angles, thereby creating perspective, or framing effects. The concept of framing was introduced in the decision-making literature by Tversky and Kahneman (1981), showing for example how an outcome framed in terms of "lives saved" led to different risk preferences than the same problem framed in terms of "lives lost." The early framing studies were used to demonstrate violations of a central axiom of rationality, called description invariance. Different ways of describing the same problem should not lead to different inferences or conclusions, as long as the facts remain unchanged. More recently, McKenzie and colleagues (McKenzie & Nelson, 2003; Sher & McKenzie, in press) have suggested that different frames are not informationally equivalent, in fact they "leak" information about expectations and reference values. When we talk about lives lost, we are talking about negative outcomes, and are making an implicit comparison to a state of affairs where fewer lives are lost. By talking about lives saved, we imply a downward comparison to situations where no or fewer lives are saved.

Similarly, we can describe the same quantity in different frames, or from opposite perspectives, by choosing a lower or an upper bound. Assume that an unspecified number between 200 000 and 300 000 people were killed by the tsunami. A newspaper headline "More than 200 000 flood victims" highlights the magnitude of the disaster. "Less than 300 000 victims" would sound less appropriate, as if the speaker is trying to reduce the dimensions of the catastrophe. There may also be a difference between inclusive and exclusive phrases in this respect, for instance an ad saying: "Cars from $20 000" will probably attract more customers than "Cars above $20 000." The studies reported in the present paper concern situations where people are instructed to provide single lower or single upper bounds. In everyday situations, people are often free to choose whether they will produce lower or upper bounds, or both, and thereby set the stage for framing effects. Framing implications of upper and lower bound estimates have been the subject of a separate study (Teigen et al., in press).

DOI: 10.1002/bdm

In contrast with most other studies of confidence intervals, we did not assign a level of confidence to the interval estimates. When such values are included in the instructions, by asking for instance for interval bounds that would in ''nine out of ten cases contain the true value,'' one might suggest inclusive rather than exclusive values; one might even think that still more extreme values are possible. Such instructions may lead to tight intervals, as respondents feel they are asked to provide plausible, ''realistic'' boundary estimates. To counteract this bias, the present research suggests a more deliberate use of single-bound exclusive estimates: The project will be finished before .... It will cost less than .... (or, depending on focus of interest: The project will take more than .... months. It will cost more than ....). Moreover, to reduce or to avoid overconfidence, we suggest that participants themselves should estimate their degree of confidence, after the interval bounds have been generated, rather than having a high level of confidence assigned to them by the experimenter in beforehand (cf. Winman et al., 2004). For one-sided intervals we found that hit rates exceeded confidence estimates, producing underconfidence in both the two first experiments. In the next two experiments, participants did not produce confidence estimates, so we do not know the relation between confidence and interval width for variable quantities. This could be a subject for further studies. However, we do not believe that confidence plays a decisive role in determining interval bounds. Previous studies (Teigen & Jørgensen, 2005) have shown that width of interval remains largely the same under very different levels of confidence (for instance 70% vs. 90% confidence, or ''I believe that ...'' vs. ''I am certain that ...'').

The results from the present studies do not lead to a specific recommendation as to whether inclusive or exclusive estimates should be preferred in interval judgments. With the two-point method (lower and upper bounds), choice of terms did not seem to make a difference (Experiment 1). With single-bound estimates, choice of terms matters, to a considerably greater extent than realized by most listeners (Experiment 5). Exclusive estimates have the advantage of being easier to produce and were associated with a higher degree of confidence (Experiments 1 and 2). But they may also be less informative, at least for estimates of specific target values. Our results also show that they show a greater amount of variability. For instance, the same product can be estimated to cost less than \$100, or less than \$1000, whereas estimates of maximum price will not vary to the same extent. It can also be argued that estimates of $T <$ \$100 and $T <$ \$1000 are not necessarily in conflict, as the former expression actually implies the latter. It follows that exclusive estimates can be influenced by factors that have little to do with the perceived target values. If I estimate the price of a car to be ''below \$10 000,'' the choice of \$10 000 as my reference point could be motivated by the price of another car (priced \$10 000, or above) or my cash balance (which happens to be \$10 000), as much as the car's actual worth. The same car could under other circumstances be estimated to cost ''below \$15 000,'' for instance in conversation with a prospective buyer who has \$15 000 as his upper price limit. Inclusive estimates, as for instance maximum prices, should be less ambiguous by reflecting the perceived price range rather than the buyer's financial situation.

REFERENCES

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

Bazerman, M. H. (1994). *Judgment in managerial decision making* (3rd ed.). N.Y.: Wiley.

Bolger, F., & Harvey, N. (1995). Judging the probability that the next point in an observed time-series will be below, or above, a given value. *Journal of Forecasting, 14*, 597–607.

Connolly, T., & Dean, D. (1997). Decomposed versus holistic estimates of effort required for software writing tasks. *Management Science, 43*, 1029–1045.

Gigerenzer, G., & Hoffrage, U. (1999). Overcoming difficulties in Bayesian reasoning: Reply to Lewis and Keren (1999) and Mellers and McGraw (1999). *Psychological Review, 106*, 425–430.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 106*, 180–209.

DOI: 10.1002/bdm

Griffin, D., & Brenner, L. (2004). Perspectives on probability judgment calibration. In D. J. Koehler, & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 177–199). Oxford: Blackwell.
Griffin, D. W., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology, 24*, 411–435.
Hacking, I. (1975). *The emergence of probability.* Cambridge: Cambridge University Press.
Jørgensen, M., Teigen, K. H., & Moløkken, K. (2004). Better sure than safe? Overconfidence in judgment based software development effort prediction intervals. *Journal of Systems and Software, 70*, 79–93.
Juslin, P., Wennerholm, P., & Olsson, H. (1999). Format dependence in subjective probability calibration. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1038–1052.
Kahneman, D., & Tversky, A. (1982). Variants of uncertainty. *Cognition*, *11*, 143–157.
Kerzner, H. (2001). *Project management: A systems approach to planning, scheduling, and controlling*. New York: Wiley.
Klayman, J., Soll, J. B., Gonzáles-Vallejo, C., & Barlas, S. (1999). Overconfidence: It depends on how, what, and whom you ask. *Organizational Behavior and Human Decision Processes, 79*, 216–247.
Liberman, V. (2004). Local and global judgments of confidence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 729–732.
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. 306–334). Cambridge: Cambridge University Press.
McKenzie, C. R. M., & Nelson, J. D. (2003). What a speaker's choice of frame reveals: Reference points, frame selection, and framing effects. *Psychonomic Bulletin & Review, 10*, 596–602.
Moder, J. J., Phillips, C. R., & Davis, E. W. (1995). *Project management with CPM, PERT and precedence diagramming*. Wisconsin: Blitz Publishing Company.
Morgan, M. G., & Henrion, M. (1990). *Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy research.* Cambridge: Cambridge University Press.
Mussweiler, T., & Strack, F. (1999). Hypothesis-consistent testing and semantic priming in the anchoring paradigm: A selective accessibility model. *Journal of Experimental Social Psychology, 35*, 136–164.
Sher, S., & McKenzie, C. R. M. (in press). Information leakage from logically equivalent frames. *Cognition*.
Sieck, W. R., & Arkes, H. R. (2005). The recalcitrance of overconfidence and its contribution to decision aid neglect. *Journal of Behavioral Decision Making, 18*, 29–54.
Sniezek, J. A., & Buckley, T. (1991). Confidence depends on level of aggregation, *Journal of Behavioral Decision Making, 4*, 263–272.
Soll, J. B., & Klayman, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 299–314.
Teigen, K. H. (2006). More than X is a lot: Pragmatic implicatures of one-sided uncertainty intervals. Unpublished manuscript, University of Oslo.
Teigen, K. H., Halberg, A. M., & Fostervold, K. I. (in press). Single-limit estimates as reference points. *Applied Cognitive Psychology.*
Teigen, K. H., & Jørgensen, M. (2005). When 90% confidence intervals are 50% certain: On the credibility of credible intervals. *Applied Cognitive Psychology, 19*, 455–475.
Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*, 453–458.
Winman, A., Hansson, P., & Juslin, P. (2004). Subjective probability intervals: How to reduce overconfidence by interval evaluation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 1167–1175.
Yaniv, I., & Foster, D. P. (1995). Graininess of judgment under uncertainty: An informativeness-accuracy trade-off. *Journal of Experimental Psychology: General, 124*, 424–432.
Yaniv, I., & Foster, D. P. (1997). Precision and accuracy of judgmental estimation. *Journal of Behavioral Decision Making, 10*, 21–32.
Yaniv, I., & Schul, Y. (1997). Elimination and inclusion procedures in judgment. *Journal of Behavioral Decision Making, 10*, 211–220.
Yaniv, I., Schul, Y., Raphaelli-Hirsch, R., & Maoz, I. (2002). Inclusive and exclusive modes of thinking: Studies of prediction, preference, and social perception during parliamentary elections. *Journal of Experimental Social Psychology, 38*, 352–367.

Authors' biographies:

Karl Halvor Teigen is a professor in general psychology at the University of Oslo, and an adjunct professor at the University of Tromsø. His interests include subjective probability judgments, social cognition, and the history of psychology.

DOI: 10.1002/bdm

Anne-Marie Halberg is a research fellow at the University of Oslo. Her interests include subjective probability judgments, psychological research methods, and measuring and analyzing constructs in organizational psychology.

Knut Inge Fostervold is an associate professor at Lillehammer University College. His interests include cognition, visual perception, judgment and decision-making, and work and organizational psychology.

Authors' Addresses:

Karl Halvor Teigen and **Anne-Marie Halberg**, Department of Psychology, P.O.B. 1094 Blindern, University of Oslo, NO-0317 Oslo, Norway.

Knut Inge Fostervold, Lillehammer University College, Norway.

DOI: 10.1002/bdm