International Journal of Managing Projects in Bu

Entirely possible overruns: How people think and talk about probabilistic cost estimates

Journal:	*International Journal of Managing Projects in Business*
Manuscript ID	IJMPB-06-2018-0114.R1
Manuscript Type:	Research Paper
Keywords:	Uncertainty, Probability / statistics, Estimation, customer perspective, effectiveness of communication, Knowledge representation

Abstract

Purpose – The purpose of this paper is to examine people's understanding and evaluation of uncertainty intervals produced by experts as part of a quality assurance procedure of large public projects.

Methodology – Three samples of educated participants (employees in a large construction company, students attending courses in project management and judgment and decision making, and judges of district and appeal courts) answered questionnaires about cost estimates of a highway construction project, presented as a probability distribution.

Findings – The studies demonstrated additivity neglect of probabilities that are graphically displayed. People's evaluations of the accuracy of interval estimates revealed a boundary (a "cliff") effect, with a sharp drop in accuracy ratings for outcomes above an arbitrary maximum. Several common verbal phrases (what *can* happen, is *entirely possible*, and *not surprising*) which might seem to indicate expected outcomes, were regularly used to describe unlikely values near or at the top of the distribution (an extremity effect)

Limitations – All judgments concerned a single case and were made by participants who were not stakeholders in this specific project. Further studies should compare judgments aided by a graph with conditions where the graph is changed or absent.

Practical implications – Experts and project managers cannot assume that readers of cost estimates understand a well-defined uncertainty interval as intended. They should also be aware of effects obtained by describing uncertain estimates in words.

Originality/value – The studies show how inconsistencies in judgment affect the understanding and evaluation of uncertainty intervals by well-informed and educated samples tested in a maximally transparent situation. Readers of cost estimates seem to believe that precise estimates are feasible and yet that costs are usually underestimated.

1. Introduction

Plans for large construction projects in the public sector in most countries undergo a detailed analysis of potential costs and benefits before being accepted by the political authorities (Samset et al., 2006). A central part of the evaluation process includes estimates of costs of selected alternatives. Such estimates will always be surrounded by considerable uncertainty and can accordingly be formulated as an uncertainty interval, or range, ideally in the shape of a probability distribution of the full span of potential outcomes. For instance, public projects with an assumed cost of approximately 100 million dollars or more will in Norway have to undergo a quality assurance procedure arriving at estimated minimum, maximum and expected costs corresponding to P15, P85 and P50 in a cumulative distribution of cost estimates. The performance of projects subject to this procedure has been discussed elsewhere (Welde, 2017). In this paper we make a first attempt to explore how external readers, who are not a part of the estimation or decision-making process, understand, perceive, and discuss such estimates. This issue is an important one, as costs occupy a central part of the public debate surrounding large investments, both at the planning stage and later when the actual outcomes can be compared with the original expectations. Those who communicate and comment upon the costs of such projects are often not themselves experts on cost estimation, but rather journalists, political analysts, or lay people concerned with the results of public expenditures (the use of "tax payers' money"). Moreover, it is known from the research literature on judgment and decision making that lay conceptions of probability and uncertainty often deviate from formal models in ways that lead to misunderstandings and unwarranted beliefs (Griffin et al., 2002; Hardman, 2009; Koehler & Harvey, 2004). Also, attempts to simplify and make formal concepts more accessible by use of everyday language, may turn out to carry implications other that intended. In the present study, we have selected the following issues for a closer examination.

- How do people assess and interpret interval estimates?
- What are their perceptions of the probabilities involved?
- Which outcomes are typically described by selected verbal phrases?

We report three studies with well-educated respondents with different, but relevant occupational background (employees in a major construction company, students following courses in decision making and project planning, and court judges), who were asked to perform judgments concerning a realistic road construction scenario, as illustrated in Figure

1.[1] All questionnaires included a graph showing the complete probability distribution of cost estimates. This would allow participants to make well-informed judgments and might be helpful in making them overcome some widespread biases in lay probabilistic thinking, on the three selected issues mentioned above.

2. Conceptual and empirical background

Empirical studies of judgment and decision-making have revealed a number of characteristics that distinguishes subjective assessments from those that are based on objective models. They are also more individual and flexible, sometimes leading to biased or inconsistent evaluations and predictions. Of special relevance for the present concerns are (1) people's views of interval boundaries, (2) the way they estimate probabilities, and (3) the language used to indicate expectancies and uncertainties. Below we examine the background for three specific judgmental effects within these areas, in turn, namely the *boundary effect*, *additivity neglect*, and the *extremity effect* in verbal probabilities.

2. 1. *Popular conceptions of interval estimates: The boundary effect*

Subjective interval estimates of uncertain past and future facts are usually too narrow. Confidence intervals that are intended to capture 90% of all possible outcomes typically miss about half of them (Connolly and Dean, 1997; Soll and Clayman, 2004). This has been regarded as a demonstration of overconfidence, but is more aptly termed over-precision (Moore and Healey, 2008). Overly narrow ranges can be given several plausible explanations (Moore et al., 2016). Estimates are supposed to be informative (Yaniv and Foster, 1995), and knowledgeable experts producing such estimates are supposed to be precise (Løhre and Teigen, 2017). In addition, people typically search for plausible (imaginable), rather than implausible minimum and maximum values, which prevent them from suggesting ranges that can incorporate surprises, especially in domains that are inadequately mapped and poorly understood, as is frequently the case with complex projects (Atkinson et al., 2006).

The width of an uncertainty interval is formally dependent upon the required level of confidence. To capture all possible outcomes with 99 or 100 percent certainty is next to impossible, except with ranges too wide to be perceived as meaningful. Unfortunately,

[1] Estimates of cost and effort of single component items typically form positively skewed curve, with longer tails for overruns than underruns. However, the sum of several such curves will yield a combined probability density function for total costs that is approximately symmetrical, as shown by Figure 1 (Halkjelsvik and Jørgensen, 2018).

people's intuitions about ranges seem unrelated to degree of confidence. People typically produce identical intervals regardless of assigned level of confidence (Langnickel and Zeisberger, 2016; Teigen and Jørgensen, 2005), oblivious to the fact that a 90% confidence interval must be considerably larger (perhaps twice as large) than a 60% interval. When asked in retrospect how confident they are in their range estimates, people often report a lower degree of confidence even if they initially were required to be 90% sure. Thus, they appear better at evaluating intervals than producing them (Martin et al., 2012). As a result, overconfidence in interval judgments can be reduced by asking experts for intervals with low rather than high degrees of certainty, or by collecting their estimates of certainty after the intervals have been produced (Speirs-Bridge et al., 2010; Teigen and Jørgensen, 2005).

In most studies of the accuracy of range estimates, participants' confidence judgments have been compared to empirical hit rates. This requires statistical information about the actual frequencies of repeated events, or multiple questions about facts that are similar enough to be regarded as a set (e.g., the population of different European capitals; the set of temperatures measured in a specific month). Less is known about how subjective range judgments are related to a theoretically derived probability distribution for a specific project, as in the case in quality assurance estimates. Do people correctly anticipate the calculated min-max interval in the report, or do they expect a narrower range than offered by the independent team of experts?

While narrow range estimates are perceived as more informative (Yaniv and Foster, 1995), and appear more certain (Løhre and Teigen, 2017) at the time of their proposal, they run a greater risk of missing the actual outcome and hence be considered "wrong" later on, when the results are known. A recent study of climate related predictions showed that people were willing to consider all outcomes that happened to fall inside a wide 90% prediction interval as having been equally well predicted. Outcomes outside the prediction interval indicated that the prediction had been wrong, this time in proportion to the distance between the outcome value and the interval boundaries (Teigen et al., 2018). A narrow, but compatible 60% interval, was in contrast viewed as producing wrong predictions. Thus, the judged correctness of the estimates depended on the actual placement of the boundary values, rather than the probabilities involved. This study did not present the shape of the probability distribution. Some people may assume that all alternatives inside an uncertainty interval are equally probable, as indicated in a study by Dieckman et al. (2015). To better illustrate the nature of a probability distribution, as well as the arbitrariness of boundary values, we introduced in the present study a bell-shaped graph with selected boundary values marked as a visual aid for participants' judgments.

2.2 *Probability estimates are frequently exaggerated: Additivity neglect*

When people are asked to attach numerical probabilities to events they seem to do so by weighing evidence in support of a target outcome against evidence that seems to be non-supportive (Tversky and Koehler, 1994), often failing to take less prominent aspects of the situation into account, like base rates, or the number of alternative outcomes. As a result, probabilities are frequently overestimated. For instance, students overestimated the chance of next Sunday to be hotter (vs. not hotter) than any other day next week (Fox and Rottenstreich, 2003), because this way of phrasing the question made them focus on two rather than seven alternative outcomes. Studies of additivity show that the sum of probabilities assigned to a set of mutually exclusive outcome often exceeds 100% (Redelmeier et al., 1995), even in cases where all alternatives are judged by the same individual (Fox and Tversky, 1998; Riege and Teigen, 2014; Teigen, 1983). When people are confronted with their inconsistent responses they typically explain that they had evaluated each separate outcome independently (Riege and Teigen, 2014). They might accordingly be overconfident with respect to "hits" within the defined uncertainty interval, and also overestimate the chances of "misses" (outcomes above the upper and below the lower boundary), if asked separate questions about expected values, overruns and underruns. For instance, they may regard costs in the $100-200 million interval as 80% likely, and yet think that even higher costs are more than 10 or 20% likely, in disregard of the 100% rule. We are not aware of studies that have systematically explored the additivity of such partitions of the outcome space.

Relatedly, people occasionally fail to distinguish between the probability of a point prediction and the prediction of an interval. Students who were asked to estimate the probability of heights in student population gave similar estimates for exact heights as for intervals (Teigen, 1974, Experiment 2). For instance, the probability of meeting a female student exactly 157 cm tall was estimated to be 0.27 (clearly an exaggerated value); the probability associated with the interval of 155-160 cm was also 0.27; and the estimate of *any* height below 160 cm was almost the same, namely 0.26. Again, one might expect that a visually presented probability distribution would help respondents overcome this kind of neglect of interval size.

2.3 *Verbal probabilities and the extremity effect*

In addition to numerical range and probability estimates, people describe and discuss uncertain outcomes verbally in the discourse of daily language. They say what they *expect*, *hope*, *believe* or *doubt* will happen, which estimates they consider *likely* or *unlikely*, what has *a chance* of happening, what is *entirely possible*, and what is *almost certain*. There is a long research tradition of trying to establish numerical translations of such phrases. For instance, Lichtenstein and Newman (1967) asked employees in a large company to place phrases like *good chance* and *rather unlikely* on a probability scale from 0 to 1. *Good chance* received a mean score of .75, and *rather unlikely* corresponded, on average, to a probability of .25, but individual scores varied widely, from .25 to .96 in the first case, and from .01 to .60 in the second. The conclusion from this research is that probability words are numerically vague and fuzzy (Budescu and Wallsten, 1995). They are, however, more precise and definite than numbers in a different respect, namely by virtue of being either affirmations, indicating a target outcome's prospect of occurrence, or negations, asking a listener to consider it might not occur. The same option can be described as having "a possibility" (positive) of success, or as being "quite uncertain" (negative). Thus, term selection reveals the speaker's attitudes, preferences or recommendations, in addition to indicating, perhaps more vaguely, the probabilities involved (Honda and Yamagishi, 2016; Teigen and Brun, 1995).

Attempts have been made to construct verbal scales corresponding to numerical probabilities in several domains, like in climate research (Mastrandrea et al., 2010), medicine (Mazur and Hickam, 1991), medication risks (EEC, 1998), and military intelligence (Barnes, 2015). Such attempts are motivated partly from a wish to make probability assessments more generally accessible and understandable, and partly to promote precision in the use of language. But it is hard to legislate natural language, and formal definitions of selected phrases may conflict with natural usage of the same expressions (Berry et al., 2003; Budescu et al., 2012). A recent line of research has shown that the "translation approach" (which probabilities correspond to *a good chance*) should be complemented with a "which outcome approach" (which outcomes are characterized as having *a good chance*) (Juanchich et al., 2013; Teigen et al., 2014). For instance, *unlikely* is commonly translated as corresponding to a 10-30% probability (Mastrandrea et al., 2010; Theil, 2002). However, it turns out that outcomes described as *unlikely* are those that have never (hitherto) occurred, in other words values with a minimal (close to zero) probability of occurrence, mostly values beyond the top of the distribution. This phenomenon has been labelled *an extremity effect* (Jenkins et al., 2018; Teigen et al., 2013). Statements about what is *possible*, or what *can*, *could*, and *may* occur will typically describe top values that are a bit more realistic than the unlikely ones, but which still, due to their extremeness, have a relatively low probability of occurrence (Teigen, Filkuková and Hohle, 2018). As these terms are often “translated” to denote a medium probability (around 50%), a lack of understanding of the extremity effect can lead to serious miscommunications.

In the present investigation, we use the pragmatic “which outcome” (sentence completion) approach to study perceived usage of selected colloquial phrases, including some phrases (*perhaps*, *entirely possible*, *not likely*, and *not surprising*) that have not been studied by this methodology before. Specifically, we look for evidence of an extremity effect in a context of cost estimates presented as a complete probability distribution with defined minimum and maximum values.

To sum up: The present studies were designed to examine selected judgmental biases in evaluating uncertainty of cost estimates, with the main focus on *additivity* neglect of probabilities, boundary effects in the evaluation of intervals, and the extremity effect in usage of verbal phrases. We hypothesize that all these effects are observable in the responses to questionnaires by readers not involved in the estimation process, but might be partly neutralized or reduced by a presentation format that makes the full probability distribution available for visual inspection.

3. **Study 1: Judgments by professionals**

This study was conducted to elicit judgments from people who through their work were familiar with planning, discussions and management of road construction projects. We assumed that these people have considerable knowledge of the uncertainties involved, and the way such uncertainties can be described (including verbal probability phrases), but were not necessarily conversant with probability calculations and the formal properties of a probability distribution of outcomes, thus they might be aided by a graphical presentation of the chances involved. They will also have opinions, based on experience, about successful and unsuccessful cost estimates, that might influence their responses to the case presented to them.

3.1 *Method*

Participants. An online questionnaire was distributed to employees in a regional division of a large construction company in Norway. Altogether 198 employees opened the questionnaire, but only 48 (25%) completed it. Of these, 43 were men and 5 women, with a median age of 45 years.

Questionnaire. The questionnaires contained a brief description about the quality assurance procedure of large public projects in Norway. They were told that an independent expert team calculated expected (most likely) costs along with a minimum and maximum estimate. The meaning of these values was briefly described and illustrated with a symmetric bell-shaped curve for a real highway project where P15, P50 and P85 were clearly marked, as shown by the graph in Figure 1. Observe that minimum and maximum did not refer to the lowest and highest value in the complete distribution, but to P15 and P85, that is, the low and high boundaries of the 70% interval. Expected costs for the recommended alternative was stated to be NOK 1100 million (approximately USD 135 million), whereas the estimates corresponding to P15 and P85 were not revealed. In the original document these values were estimated to be NOK 750 million (USD 90 million) and NOK 1450 million (USD 180 million), respectively.

<Insert Figure 1 about here>

The graph was followed by 11 questions, of which Questions 1-6 and 8 are relevant to the present concerns and will be reported here (translated from Norwegian[2]).

Q1 What do you think were estimated as minimum and maximum values by the expert team?

Q2 Imagine that you had to describe in informal language what this highway project would cost including the uncertainties involved. Formulate two-three sentences without numerical probabilities (instead use words like “good chance”, “small chance”, “possible”, “uncertain” and so on).

Q3 Imagine an informal conversation between colleagues who have read the project report. Fill in an amount of costs that makes sense and appears natural in this context.

Altogether 9 statements were presented in random order: “The project will cost more than / less than / will probably[3] cost / not probable it will cost / improbable it will cost / can (could)[4] cost / will perhaps cost / entirely possible it will cost / I would not be surprised if it costs …. million”.

---

[2] For the complete questionnaire, see Teigen, Andersen and Alnes (2018).

[3] The Norwegian term «sannsynlig» can be translated with English *probably* or *likely*. Similarly, *unlikely* and *improbable* have to be translated into Norwegian with the same term, “usannsynlig”.

[4] The Norwegian term was «kan», which in the present context corresponds to English *can*, *could*, or *may*.

Q4 Which three statements (of these) would you use in a conversation with the project owner

Q5 Imagine that a journalist listens to the conversation. Which three statements do you think he will write down for potential use in his newspaper report?

Q6 Imagine that the team responsible for making the estimates were asked about their probabilities (as numbers between 0 and 100%) – what do you think they would say?

As probabilities of exact point estimates do not make much sense this question was qualified as referring to approximate values (±50 million), i.e., most likely cost estimate (±50 million), minimum estimate (±50 million), and maximum estimate (±50 million). Values for minimum and maximum matched those provided by the individual respondent (piped from their answers to Q1).

Q8 Imagine that the actual costs turned out to be ….. million. How would you rate the original estimates (on seven-point scales from 1: Completely wrong, to 7: Completely correct).

This question was repeated three times. The first time actual costs were set equal to the maximum suggested by the individual respondents (piped from their answers to Q1), the second time with NOK 100 million *lower*, and the third time with NOK 100 million *higher* than the respondents' own maximum estimates. Thus, respondents had to decide whether an outcome at the boundary of their own uncertainty interval had been correctly predicted, compared to outcomes inside or outside of this interval.

3.2 *Results*

Width of uncertainty intervals and boundary effects. In Q1, Participants underestimated the width of the uncertainty interval, as predicted. The original quality assurance report had placed P15 at 750 million and P85 at 1450 million, defining a 700 million uncertainty interval, whereas 91% of respondents proposed a narrower range, with M_{P15} = 954 million (Median = 945 million) and M_{P85} = 1343 million (Median = 1300 million), yielding an uncertainty interval about half as wide. Thus over-precision (the tendency to produce too narrow ranges), previously found for assessors' own confidence intervals (Moore et al., 2016), can also be demonstrated for their estimates of others' ranges, and for probability intervals supposed to cover only 70% of the total distribution.

In response to Q8, participants saw the estimate as generally correct (M_{P85} = 4.34) when actual outcomes corresponded to the proposed maximum. It was seen as slightly more correct ($M_{P85\text{-}100}$ = 4.82) when actual outcome is 100 million lower and accordingly inside the interval, but clearly wrong when actual outcome is 100 million higher than the maximum and thus outside of the predicted range ($M_{P85+100} = 3.07$) (see also Figure 2, first bars in each set). An overall ANOVA for repeated samples indicate a difference among the means, $F(2, 86) = 15.23$, $p < .001$; separate comparisons show that the first two means are not significantly different, whereas the third mean is significantly different from each of the two first ones, $t(43) = 4.39$ and 4.24, $p < .001$, indicating a boundary(or "cliff") effect.

Probability judgments. Mean probabilities of three selected outcomes, in answer to Q6, are displayed in the first column of Table 1. These probabilities are far too large to describe three rather small slices of the outcome distribution, and have perhaps been mixed up with interval estimates or even with cumulative probabilities by some participants, despite explicit instructions to estimate three separate values, surrounded by narrow margins. So instead of producing three probabilities of 5-15% (as might be inferred by a visual inspection of the graph) they suggested probabilities adding up to more than 100%, showing additivity neglect. Observe also that outcomes around the maximum value are considered much more likely than outcomes around the minimum, despite the symmetry of these values displayed in the graph.

<Insert Table 1 about here>

Verbal probabilities and extremity effects. When participants were asked to formulate expectations in words (Q2), they tried to balance chances against uncertainties, focusing more strongly on expenses than on potential savings. Illustrative examples:

- "Small chances of road project becoming cheaper than supposed, great chances of considerable additional expenses".
- "The road will most likely cost between 1265 and 935 million but a good chance of higher or lower costs".

As indicated by the first statement, not all participants accepted the premise of a symmetrical distribution, which they had explicitly been asked to describe. The second statement suggests a willingness to regard each of several exclusive outcomes as likely at the same time, demonstrating a kind of additivity neglect with verbal phrases.

When asked to complete verbal probability statements with appropriate costs, participants suggested, as expected, low amounts (equal to or smaller than P15) in *more than*-statements, and large amounts (equal to or higher than P85) in *less than*-statements, as displayed in the left panel of Table 2. Statements with *probably* (*likely*) were predictably completed with an intermediate amount, corresponding to P50 (1100 million). For the remaining expressions, various extremity effects could be observed. Costs that are characterized as *not likely*, or *unlikely*, belong to the tails of the distribution, they are either very high or very low. Perhaps more interesting, costs that *can* (could) occur, those that are *entirely possible*, or claimed to be *unsurprising* are not among the likely ones, but belong mostly to the high end of the distribution.

More than and *probable* are preferred in conversation with an employer (Q4), whereas it was supposed that the journalist in Q5 would rather make a note of the more informal phrases: *can* cost, *entirely possible*, and *not surprising*, all phrases that suggest that costs will run high, which probably is regarded as more newsworthy.

<Insert Table 2 about here>

4. **Study 2: Judgments by students**

Participants in Study 1, who presumably knew project work and cost estimates from their own work experience, were biased towards thinking that costs would be higher than assumed, even when explicitly asked to relate to the distribution displayed in the graph rather than expressing their personal opinions. In Study 2 the same task was presented to students following courses in judgment and decision making and project management at bachelor and master levels, for whom it may be natural to have a more theoretical than practical approach to uncertainty intervals. In Study 2 the "difficult" (and apparently misunderstood) question about the probability of point estimates (Q6) was replaced by an easier question about the probability of intervals that could be directly read from the graph.

4.1 *Method*

Participants. The link to an online questionnaire was distributed to university students attending classes at two different Norwegian universities. Students in Group 1 attended a master level course in project management. They received the questionnaire in English, which was opened by 92 students and completed by 51 (55%), 42 men and 9 women, with a median age of 24 years. Students in Group 2 were 36 bachelor students (9 men and 27 years, median age 21 years) who participated a class in the psychology of judgment and decision making. All students in this group completed the questionnaire.

Questionnaire. The students received the same questions as in Study 1, with the exception that Q2 (informal explanation of expectations and uncertainties in one's own words) was omitted. In addition, Q6 was changed into a question about the probability of intervals rather than points, namely the probability of costs equal to or less than minimum, the probability of costs between supposed min and max values, and the probability of costs equal to or above the supposed maximum. Since these values were defined as P15, P50 and P85 these three probabilities could be copied directly from the graph as corresponding to 15%, 70%, and 15%, respectively. As an extra reminder, half of the participants in Group 1 received the graph in Figure 1 for the second time immediately above Q6.

4.2 *Results*

Width of uncertainty intervals and boundary effects. A majority (65%) of project students in Group 1 underestimated the width of the uncertainty interval, against only 47% of JDM students in Group 2. The range proposed by Group 1, with M_{P15} = 939 million (Median = 917 million) and M_{P85} = 1482 million (Median = 1300 million), was similar to the range proposed by professionals in Study 1. A one-way ANOVA where Study 1 and the two groups of Study 2 are compared, gives $F(2, 119) = 6.265$, $p = .003$. Post hoc tests (LSD) reveal that the intervals of Group 2 were significantly wider than the intervals for professionals ($p = .001$) and Group 1 ($p = .014$), whereas ranges produced by professionals in Study 1 and students in Group 1of Study 2 were not significantly different from each other. The surprisingly wide intervals offered by the bachelor students in Group 2 was most likely inspired by a recent lecture about overconfidence and the narrowness of prediction intervals attended by students in this group. Apparently, students sometimes learn from what they are taught in class!

Correctness ratings for three sets of outcomes, 100 million below max (P85), equal to max, and 100 million above max are shown in Figure 1. The three first bars in each set give mean scores for Study 1 and the two groups of Study 2. It is apparent from a visual inspection of the figure that the ratings of the first and second set are more similar ($M_{diff} = 0.63$). than the second and third set ($M_{diff} = 1.19$). A 2 x 3 mixed ANOVA reveals a significant effect of these difference scores, $F(1, 125) = 6.499$, $p = .012$, but no effect of group and no significant interaction. Thus, outcomes just *at* the interval boundaries are viewed as having been predicted tolerably well, only slightly less accurately than outcomes inside the boundaries, and much more accurately than outcomes exceeding the upper boundary.

<Insert Figure 2 about here>

Probability judgments. Participants in this study were given an “easy” question about probabilities that could be answered by simply copying the percentages attached to the figure. But only a minority did so. Most participants were non-additive and produced three probability estimates that added up to more than 100%. Mean estimates are displayed in the right panel of Table 3, showing that they judged the probability of outcomes in the right tail of the distribution (above P85) to be almost twice the probabilities in the left tail (below P15), despite the symmetry displayed in the graph. Participants who saw the graph two times, the second time adjacent to the question about probabilities, did no better than the others.

Verbal probabilities. Median costs corresponding to selected phrases and percentages of participants choosing numbers in the tails of the distribution are shown in the right panel of Table 2. The choices agree well with those made by employees in Study 1. Except for “likely”, which is believed to describe a value in the middle range, all other phrases are expected to characterize either high or low costs. The distributions for *unlikely* and *not likely* are bimodal. *Can* (could) cost, *entirely possible*, *not surprising*, and to some extent *perhaps*, suggest high rather than low amounts of costs, in line with the extremity effect. The respondents agreed that a journalist would prefer statements about costs that *can* occur, are *not surprising,* and *entirely possible,* whereas the more neutral statements *more than, less than* and *probable* were preferable in conversations with the project owner. Preferences for terms in Study 1 and 2 are combined in Figure 3.

<Insert Figure 3 about here>

5. **Study 3: Judgments by judges**

Studies 1 and 2 gave some evidence for a boundary effect as well as for an extremity effect in both professional and student samples. In Study 3 we extended the investigation to judges, who are perhaps less familiar with project management, but presumably have considerable experience in distinguishing between correct and incorrect statements and in the interpretation of verbal phrases. As in the other studies, they received cost predictions accompanied by a visual representation of the probability distribution, but to highlight the boundary effect, correctness ratings for outcomes inside or outside of the interval were performed by different participants in a between-subjects design. We predicted that outcomes equal to the interval maximum would be rated more similar to lower outcomes, than to higher outcomes. As for verbal phrases, *entirely possible* was singled out for further investigation. This phrase seems, on one hand, to indicate an expected occurrence, but was in Study 1 and 2 primarily used in statements about high costs. We examined in the present study whether this association is bidirectional. Will statements about high (maximum) costs be more easily described as being *entirely possible*, than as having, for instance, *a low probability* of occurrence?

5.1 *Method*

Participants. Two questions pertinent to the present issues were embedded in a large online questionnaire distributed to Norwegian judges participating in an adult education course arranged by the Norwegian Courts Administration. Of 395 participants who opened the link, 356 (85%) answered the present questions, 46% women, median age 54 years. Of these, 182 reported working in district courts, 89 in appeal courts, and 5 were supreme court judges (80 held other offices or did not indicate profession). Participants were allocated to two conditions, A and B, by a semi-random procedure (choice of random numbers).

Questionnaire. The questions about cost estimates were introduced as in the preceding studies, including the graph in Figure 1. The three vertical lines were explicitly labelled Minimum 750 million, Expected 1100 million, and Maximum 1450 million, and the text explained that 15% of the probability distribution were below the minimum mark, and another 15% above the maximum.

Sentence completion. Participants in Condition A were asked to imagine an informal conversation between colleagues about the highway project. One of them says: "It is entirely possible that it will cost *750 million / 1100 million / 1450 million*". Choose the number that seems most natural in this context.

Participants in Condition B were instead asked to choose an appropriate verbal phrase in a statement about the maximum value. "It is a *low probability / probable / entirely possible*[5] that it will cost 1450 million".

We expected that the statement in A would be completed with 1450 million (in line with the extremity effect). The complementary statement in B was viewed as more ambiguous, as *entirely possible* must here compete with other phrases, and 1450 million is defined by the distribution as actually having *a low probability* of occurrence.

Accuracy ratings. Participants in Condition A rated (1-7) the correctness of cost estimates if actual costs turned out to be (1) 1350 million (100 million below maximum estimate), or (2) 1450 million (matching the maximum estimate). Condition B rated the

[5] The Norwegian terms were: Lite sannsynlig / sannsynlig / fullt mulig

correctness of cost estimates if actual costs turned out to be (1) 1450 million (matching the maximum) and (2) 1550 million (100 million above maximum). A boundary effect would result in a larger difference between the two ratings in Condition B than in Condition A.

5.2 *Results*

The judges agreed that *entirely possible* describes a high value rather than a low or intermediate one (see Table 3). The association goes both ways, as a majority also thought that a high value should be characterized as *entirely possible* rather than *probable* or having a *low probability*, despite the fact that 1450 million was defined as a maximum value, only exceeded by 15%, and had accordingly a rather low probability of occurrence.

< Insert Table 3 about here>

Participants in Condition A rated estimates as more accurate when actual costs turned out to be 1350 million (100 million below max) than 1450 million (at max), but the mean difference in accuracy ratings was rather small, $M_{\text{diff}} = 0.46$. The difference in Condition B, between 1450 million (max) and 1550 million (100 million above max), was considerably larger, $M_{\text{diff}} = 1.70$; $t(246) = 10.99$, $p < .001$ for the difference of differences. Mean ratings are displayed in Figure 2 as the fourth and fifth bar in each set, for Condition A and B, respectively.

6. General discussion

Returning to our initial questions, we see that educated groups of people experience several difficulties and dilemmas in evaluating uncertainty intervals for project costs, even when aided by a visual representation of the probability distribution. Their probability estimates appeared excessive and non-additive; arbitrarily placed interval boundaries played a decisive role in evaluating the accuracy of estimates; and a verbal expression that seemingly denote a representative outcome was used to describe extreme and hence unlikely outcomes.

Graphical displays of variability and probability have in other studies been shown to improve probabilistic judgments (e.g., in the area of health risks) and bring them more in line with formal requirements (Burkell, 2004; Joslyn et al., 2013; Lipkus and Hollands, 1999), but even if our respondents may have found the graphical illustration helpful, some problems of interpretation and judgment seem to persist. The high agreement between the diverse samples in the present studies indicates some generality of findings. Participants in the three studies differed in age and occupation, but shared presumably an interest in estimation and proper judgments. Judgmental effects observed in these samples would probably characterize evaluations done by people with less judgmental competence as well.

6.1 *Probability estimates*

Probabilities are notoriously difficult to assess and understand without the aid of graphs and calculations. Estimates provided by participants in the present studies show that difficulties persist even when a display of the results of calculations are available. We were surprised that the information provided appeared to be partly neglected, even by project students in Study 2 who had the graph plainly in view while estimating probabilities. Both these and the employees in Study 1 may have misunderstood the task (some might even have found it too easy), and thought it asked for adding opinions of their own. A large study of how people in different countries read climate predictions revealed a similar neglect (Budescu et al., 2014). Respondents in this study were provided with a table of the numerical equivalents of selected verbal probabilities, as used by authors of the IPCC reports (Mastrandrea et al., 2010). They were then asked to interpret statements from the IPCC report where these verbal phrases occurred. Instead of simply copying the corresponding numbers from the table, most respondents produced their own deviant translations.

Some deviations, or misunderstandings, of the task might stem from the blurred distinction between points and interval predictions. The probability that a road project will cost 1100 million (±50 million) might be read as the probability that "it will cost *at least* 1100 million". After all, it is not uncommon to make *at least* interpretations of numbers (Levinson, 2000; Musolino, 2004). For instance, when statisticians speak about the probability (under the null hypothesis) of a specific difference between two means, they mean the probability of a difference *at least* this high. But an *at least* reading of numbers can hardly justify probability estimates around or above the maximum as 50% or 25%, as Table 1 suggests. The respondents seemed to entertain a strong belief that costs for large projects are likely to exceed the original, perhaps too optimistic estimates. This belief is not without empirical support (Flyvbjerg, 2016; Morris and Hough, 1987; Prater, Kyrotopoulos and Ma, 2017), and is reinforced by the attention drawn to budget "cracks" of public projects exposed in the media. As a result, respondents in the present studies were drawn between the implications of the symmetrical probability distribution presented to them (the graph in Figure 1), which said that high and low estimates are expected with the same probability, and their prior beliefs about overruns as more likely than underruns. These two coextensive, conflicting pieces of evidence led to asymmetric inferences from a symmetric curve.

6.2 *Boundary effects*

The boundary effects observed in the present studies were weaker than those reported in Teigen et al.'s (2018) climate prediction studies. This could be attributed to the visual presentation of a continuous distribution extending beyond the maximum and minimum bounds. The effect might also have been attenuated by design features, as participants in the present studies were asked to produce accuracy ratings of more than one potential outcome (three in Study 1 and 2, and two in Study 3), which hence could be directly distinguished and compared, perhaps suggesting a graded judgment rather than a dichotomy between estimates inside or outside of the uncertainty interval. Still, these arbitrary boundary values led to an obvious "cliff" effect with respect to the evaluation of an estimate as right or wrong, as shown in Figure 2. Observe that participants in Study 1 and 2, did not receive the actual minimum and maximum estimates, but judged prediction accuracy relative to their own suggestions of plausible maximum values. This means that one specific outcome (e.g., of 1450 million) would be regarded as incorrectly predicted by participants suggesting narrow intervals, but not by participants who thought, or were told that the intervals were this wide (for instance judges in Study 3).

Boundary effects can be justified if the boundary values form the basis for subsequent decisions. For instance, estimated "maximum" values have been used to define the cost frame for projects approved by the authorities in Norway. This means that costs no higher than the upper boundary are acceptable, whereas higher costs are considered cracks. However, we cannot assume that our participants were familiar with this practice, and find it doubtful that it would have influenced their ratings. It is more likely that this practice and the present ratings have a common basis in a general human tendency to impose categories on continuous variables, with assimilation *within* each category and contrast *between* categories as a result (Eiser and Stroebe, 1972; Tajfel and Wilkes, 1963). Such effects have been demonstrated in several domains, both with natural and arbitrary category boundaries, for a variety of tasks (e.g., Krueger and Clement, 1994; Rothbart et al., 1997).

6.3 *Verbal probabilities*

The present studies provide new evidence for extremity effects in the usage of verbal phrases describing uncertain outcomes. While the traditional "translation approach" has tried to understand such phrases by asking people to indicate the corresponding numerical probabilities, we ask in this study *which outcome* they think a speaker has in mind. These two approaches correspond well for *probable* (or *likely*) outcomes, which in a bell-shaped distribution are to be found in the middle and are correctly associated with higher numerical probabilities than other outcomes. For many other phrases, people expect that speakers have extreme rather than intermediate outcomes in mind, regardless of the probabilities involved. Other studies (Jenkins et al., 2018; Teigen et al., 2013) indicate that *improbable* outcomes are typically outcomes above the maximum of the prediction interval, which have a near-zero probability of occurrence. The present studies yielded, in this case a bimodal distribution, as participants of Study 1 and 2 assumed that *improbable* (*unlikely*) and *not probable* (*not likely*) outcomes were either very low or very high, about equally often (see Table 2). The high frequency of very low "unlikely" outcomes in the present studies may be due to a prior belief that underruns are more exceptional than overruns, regardless of the estimated probability distribution.

Other common phrases denoting potential outcomes (*can*, *not surprising*, and *entirely possible*) were more often believed to be associated with high than low outcomes, demonstrating a different variety of the extremity effect. This usage may seem to conflict with a probabilistic interpretation of these terms and with their "positive" directionality, that makes them sound like expected rather than exceptional outcomes. A recent study (Teigen et al., 2018) showed that these two readings of the same terms can lead to miscommunication. Speakers (writers) used the word *can* to describe a maximum outcome, whereas recipients (readers) of this message assumed they spoke about one of the most likely ones. A similar speaker/listener asymmetry might occur if speakers use *entirely possible* to describe top outcomes, whereas listeners might think that an entirely possible outcome must be among those we expect will happen. However, the judges in Study 3 showed no indication of such an asymmetry, as they thought that speakers would use *entirely possible* to denote extreme costs, and also that extreme costs would be denoted by the term *entirely possible* rather than, for instance by a *low probability*. We do not know how other groups will respond to this inverted question; it is in any case reassuring to know that in this regard, judges are not easily fooled.

6.4 *Limitations and suggestions*

This study was based on estimates of a single highway construction project, and involved estimation and approval procedures in one specific (Scandinavian) scene. The samples in the first two studies were small, and particularly Study 1 had a low response rate, indicating some self-selection of participants. Yet, similarity in patterns of answers across samples suggest that the findings can claim some generality, but further studies should show whether they also apply to politicians, managers and other stakeholders involved in the actual decision-making process for such projects. All studies made use of graphical displays, adopted for maximum transparency, but the absence of a condition without graphics prevents strong conclusions to be drawn about their role, for instance whether they promote (or prevent) a reader's understanding of the arbitrariness and the probabilistic nature of interval boundaries. Different ways of presenting and explaining intervals should be compared in further studies. For instance, bell-shaped probability distributions as used in the present studies could be compared to S-shaped cumulative distributions or to descriptions without graphs (for a comparison of different graphical methods for presenting quantitative uncertainty, see Edwards et al., 2012). The meaning of probabilistic intervals would perhaps be easier to grasp if, for instance, a 90% and a 50% interval were presented together. Descriptions of P50 as "expected" (or "most likely") in a symmetric, unimodal distribution might entice people to believe that this specific outcome is actually quite likely, while the truth is that the likely or expected outcome is an outcome *within the defined range* and not a specific number. We also need research that systematically explores the relationship between probability judgments of approximate point values and wider vs. narrower partitions of the probability distribution.

7. Conclusions

The present studies show that people belonging to diverse, well-educated occupational groups experience problems when it comes to judging probabilities, interpreting the width and accuracy of range predictions, and describing chances in verbal terms. Thus, experts calculating cost estimates cannot take for granted that a well-defined and graphically illustrated probability distribution will be understood as intended by all readers. Point and interval estimates are easily mixed up, prior expectations will colour the interpretations drawn, and people's conception of uncertainty may not be fully captured by a single, "one-size-fits-all" probability distribution (Parker and Risby, 2018). Common verbal terms that are intended to give people a more intuitive grasp of the uncertainties and chances involved may be misleading, if it is not realized that they often apply to extreme outcomes and offer little guidance about probabilities. Such effects represent a challenge for experts presenting their estimates to project managers and for project managers informing employees and the public about prediction intervals of costs. Perhaps ironically, attempts to improve predictions by ranges capturing a balanced number of overruns and underruns (Welde, 2017), may not be fully appreciated by outside readers, who may continue thinking that *uncertainty* in projects means to be prepared for unpredicted (but not surprising) overruns.

The present studies were not designed to provide guidelines for estimators and project managers about how to improve their communication about uncertain project costs, except indirectly by making them aware of common non-normative interpretations of such estimates. "Assume low numeracy of a general public audience" (Spiegelhalter et al., 2011, p. 1399) may be a good piece of advice. It may be helpful to tailor communication to different target audiences depending on their presumed level of scientific literacy. On might for instance consider a categorization recommended by European Food Safety Authority (EFSA, 2018), which differentiates between audiences at an *Entry* (media, general public), an *Informed* (policy/decision-makers) and a *Technical* (scientific) level for communicating risks and uncertainties. To illustrate: Whereas a probability distribution with arbitrary cutoff points (like the one presented in Figure 1) presumably works well at the technical level, people at the informed level may find explicit information about alternative intervals helpful; but at the entry level more than one interval might be confusing. Complementary probability estimates are superfluous for the expert, but might be explicitly needed by "informed" decision makers to divert their attention from success probabilities to the (complementary) probabilities of overruns and other surprising outcomes. At the "entry" level information overload might be a problem. Also, terms denoting maximal and minimal expectations are useful to describe the range of uncertainty, but might in turn be perceived as boundaries distinguishing between 'correct' and 'incorrect' predictions. For audiences at the entry level, terms like *low* and *high* estimates might be less misleading than minimum and maximum (which taken literally exclude still lower or still higher values). Finally, verbal phrases who are often believed to simplify uncertainty communication, particularly for a lay audience, are not a substitute for numbers; they are better suited to add argumentative and evaluative flavor to a quantitative expression. For instance, it makes sense to say: "it is 20 percent chance, so it is unlikely", or: "it is 20 percent chance, so it is entirely possible", whereas the phrases *unlikely* or *entirely possible* without numbers might indicate quite different, often extreme outcomes. Their commonness in ordinary discourse should not entice communicators to believe they have a well-established probabilistic meaning.

References

Atkinson, R., Crawford, L. and Ward, S. (2006), "Fundamental uncertainties in projects and the scope of project management", *International Journal of Project Management*, Vol. 24 No. 8, pp. 687-698.

Barnes, A. (2015), "Making intelligence analysis more intelligent: Using numeric probabilities", *Intelligence and National Security,* No. 4527, pp. 1–18. https://doi.org/10.1080/02684527.2014.994955.

Berry, D.C., Raynor, D.K., Knapp, P.R. and Bersellini, E. (2003), "Patient understanding of risk: impact of EU guidelines and other risk scales for consumer medicines information", *Drug Safety,* Vol. 26, pp. 1-11.

Budescu, D.V., Por, H.H. and Broomell, S.B. (2012), "Effective communication of uncertainty in the IPCC reports", *Climatic Change,* Vol. 113 No 2, pp. 181–200.

Budescu, D.V., Por, H.H., Broomell, S.B. and Smithson, M. (2014), "The interpretation of IPCC probabilistic statements around the world", *Nature Climate Change,* Vol. 4, pp. 508-512.

Budescu, D.V. and Wallsten, T.S. (1995), "Processing linguistic probabilities: General principles and empirical evidence", in R.H. Busemeyer and D. Medin (Eds.), *The Psychology of Learning and Motivation*, Vol. 32, pp. 275–318. Academic Press, Amsterdam.

Burkell, J. (2004), "What are the chances? Evaluating risk and benefit information in consumer health materials", *Journal of Medical Library Association,* Vol. 92, pp. 200–208.

Connolly, T. and Dean, D. (1997), "Decomposed versus holistic estimates of effort required for software writing tasks", *Management Science*, Vol. 43 No. 7, pp. 1029–1045.

Dieckman, N.F., Peters, E. and Gregory, R. (2015). "At home on the range? Lay interpretations of numerical uncertainty ranges", *Risk Analysis,* Vol. 35 No. 7, pp. 1281–1295.

Edwards, J.A., Snyder, F.J., Allen, P.M., Makinson, K.A. and Hamby, D.M. (2012), "Decision making for risk management: a comparison of graphical methods for presenting quantitative uncertainty", *Risk Analysis*, Vol. 32 No. 12, pp. 2055-2070.

EEC (1998), A Guideline on the Readability of the Label and Package Leaflet of Medicinal Products for Human Use, EC Pharmaceuticals Committee, Brussels.

EFSA (2018), “Technical Report on the EFSA-Member State multilingual online survey on communication of uncertainty to different target audiences”, *EFSA Supporting Publication 2018: EN-1413*. doi: 10.2903/sp.efsa.2018.EN-1413.

Eiser, J.R. and Stroebe, W. (1972), *Categorization and Social Judgment,* Academic Press, New York.

Flyvbjerg, B. (2014), “What you should know about megaprojects, and why: An overview”, *Project Management Journal,* Vol. 45 No 2, pp. 6-19.

Fox, C. R. and Rottenstreich, Y. (2003), “Partition priming in judgment under uncertainty”, *Psychological Science,* Vol. 13, pp. 195–200.

Fox, C.R. and Tversky, A. (1998), “A belief-based account of decision under uncertainty”, *Management Science,* Vol. 44, pp. 879-895.

Gilovich, T., Griffin, D.W., and Kahneman, D., (2002), *Heuristics and Biases: The Psychology of Intuitive Judgement,* Cambridge University Press, Cambridge.

Hardman, D. (2009), *Judgment and Decision Making: Psychological Perspectives,* Wiley-Blackwell, London.

Honda, H. and Yamagishi, K. (2017), “Communicative functions of directional verbal probabilities: Speaker’s choice, listener’s inference, and reference points”, *The Quarterly Journal of Experimental Psychology*, Vol. 70, pp. 2141-2158.

Jenkins, S.C., Harris, A.J.L. and Lark, R.M. (2018), “Understanding ‘unlikely (20% likelihood)’ or ‘20% likelihood (unlikely)’ outcomes: The robustness of the extremity effect”, *Journal of Behavioral Decision Making*, DOI: 10.1002/bdm.2072

Joslyn, S., Nemec, L. and Savelli, S. (2013), “The benefits and challenges of predictive interval forecasts and verification graphics for end users”, *Weather, Climate, and Society,* Vol. 5, No. 2, pp. 133-147.

Koehler, D.J. and Harvey, N. (Eds.), The Blackwell Handbook of Judgment and Decision Making, Blackwell, Oxford

Krueger, J. and Clement, R. W. (1994) “Memory-based judgments about multiple categories: A revision and extension of Tajfel’s accentuation theory”, *Journal of Personality and Social Psychology,* Vol. 67, pp. 35-47.

Juanchich, M., Teigen, K.H. and Gourdon, A. (2013), „Top scores are possible, bottom scores are certain (and middle scores are not worth mentioning): A pragmatic view of verbal probabilities”, *Judgment and Decision Making*, Vol. 8, pp. 345-364.

Langnickel, F. and Zeisberger, S. (2016), "Do we measure overconfidence? A closer look at the interval production task, *Journal of Economic Behavior and Organization,* Vol. 128, pp. 121–133.

Levinson, S.C. (2000), *Presumptive Meanings*, MIT Press, Cambridge, MA.

Lichtenstein, S. and Newman, R. (1967), "Empirical scaling of common verbal phrases associated with numerical probabilities", *Psychonomic Science,* Vol. 9, pp. 563–564.

Lipkus, I. and Hollands, J. (1999), "The visual communication of risk", *Journal of National Cancer Institute Monographs,* Vol. 25, pp. 149–163.

Løhre, E. and Teigen, K.H. (2017), "Probabilities associated with precise and vague forecasts", *Journal of Behavioral Decision Making,* Vol. 30*,* pp. 1014-1026.

Martin, T.G., Burgman. M.A., Fidler, F. et al. (2012), "Eliciting expert knowledge in conservation science", *Conservation Biology*, Vol. 26, pp. 29–38.

Mastrandrea, M.D. et al. (2010), Guidance note for lead authors of the IPCC Fifth Assessment Report on consistent treatment of uncertainties, *IPCC*, http://www.ipcc.ch

Mazur, D. J. and Hickam, D.H. (1991), "Patients' interpretations of probability terms", *Journal of General Internal Medicine,* Vol. 6, pp. 237-240.

Moore, D.A. and Healy, P.J. (2008), "The trouble with overconfidence", *Psychological Review,* Vol. 115 No 2, pp. 502-517.

Moore, D.A., Tenney, E.R. and Haran, U. (2016), "Overprecision in judgment", in G. Wu and G. Keren (Eds.), *Handbook of Judgment and Decision Making,* pp. 182-209, Wiley, New York.

Morris, P.W.G. and Hough, G.H. (1987), The Anatomy of Major Projects: A Study of the Reality of Project Management, Wiley, New York.

Musolino, J. (2004), "The semantics and acquisition of number words: Integrating linguistic and developmental perspectives", *Cognition,* Vol. 93, pp. 1–41.

Parker, W.S. and Risbey, J.S. (2018), "False precision, surprise and improved uncertainty assessment", *Philosophical Transactions of the Royal Society A,* Vol. 373, No. 2014.0453.

Prater, J., Kirytopoulos, K. and Ma, T. (2017), "Optimism bias within the project management context", *International Journal of Managing Projects in Business,* Vol. 10, pp. 370-385.

Redelmeier, D., Koehler, D.J., Liberman, V. and Tversky, A. (1995), "Probability judgment in medicine: Discounting unspecified possibilities", *Medical Decision Making*, Vol. 15, pp. 227–230.

Riege, A.H. and Teigen, K.H. (2013), “Additivity neglect in probability estimates: Effects of numeracy and response format”, *Organizational Behavior and Human Decision Processes,* Vol. 121, pp. 41-52

Rothbart, M., Davis-Stitt, C. and Hill, J. (1997), “Effects of arbitrarily placed category boundaries on similarity judgments”, *Journal of Experimental Social Psychology,* Vol. 33, pp.122-145.

Samset, K., Volden, G.H., Olsson, N. and Kvalheim, E.V. (2016), “Governance schemes for major public investment projects: A comparative study of principles and practices in six countries”, *Concept report No. 47*, NTNU Trondheim.

Soll, J.B. and Klayman, J. (2004), “Overconfidence in interval estimates”, *Journal of Experimental Psychology: Learning Memory and Cognition,* Vol 30 No. 2, pp. 299–314.

Speirs-Bridge, A., Fidler, F., McBride, M., Flander, L., Cumming, G. and Burgman, M.A. (2010), “Reducing overconfidence in the interval judgments of experts”, *Risk Analysis,* Vol. 30, pp. 512–523.

Spiegelhalter, D., Pearson, M. and Short, I. (2011), “Visualizing uncertainty about the future”, *Science,* Vol 333, pp. 1393-1400.

Tajfel, H., and Wilkes, A.L. (1963), “Classification and quantitative judgment”, *British Journal of Social Psychology,* Vol. 54, pp. 101-114.

Teigen, K.H. (1974), “Overestimation of subjective probabilities”, *Scandinavian Journal of Psychology,* Vol. 15, pp. 56-62.

Teigen, K.H. and Filkuková, P. (2013), “Can > will: Predictions of what can happen are extreme, but believed to be probable”, *Journal of Behavioral Decision Making,* Vol. 26, pp. 68–78.

Teigen, K.H., Filkuková, P. and Hohle, S.M. (2018), „It can become 5° C warmer: The extremity effect in climate forecasts”, *Journal of Experimental Psychology: Applied*, Vol.

Teigen, K.H., Juanchich, M. and Filkuková, P. (2014), „Verbal probabilities: An alternative approach”, *Quarterly Journal of Experimental Psychology,* Vol. 67, pp. 124-146.

Teigen, K.H., Juanchich, M. and Riege, A.H. (2013), “Improbable outcomes: Infrequent or extraordinary?”, *Cognition*, Vol. 127, pp. 119-139.

Teigen, K.H., Løhre, E. and Hohle, S.M. (2018), «The boundary effect: Perceived post hoc accuracy of prediction intervals”, *Judgment and Decision Making*, Vol. 13 No. 4, pp. 309-321. http://journal.sjdm.org/17/171211/jdm171211.pdf.

Theil, M. (2002), “The role of translations of verbal into numerical probability expressions in risk management: A meta-analysis”, *Journal of Risk Research,* Vol. 5, pp. 177–186.

Tversky, A. and Koehler, D.J. (1994), “Support theory: A nonextensional representation of subjective probability”, *Psychological Review,* Vol. 101, pp. 547–567.

Welde, M. (2017), “Cost performance in large government investment projects that have been subjected to external quality assurance”, *Concept report No. 51*, NTNU, Trondheim

Yaniv, I. and Foster, D.P. (1995), ”Graininess of judgment under uncertainty: An accuracy-informativeness trade-off”, *Journal of Experimental Psychology: General,* Vol. 124 No 4, pp. 424-432.

Table 1. *Mean probability estimates for costs corresponding to three approximate point values (Study 1) and three segments of the outcome distribution (Study 2).*

Study 1		Study 2			
Employees			Project students (Group 1)		JDM students (Group 2)
Point estimates		Intervals	With graph	Without graph	
P15 (±50 million)	28.0%	≤ P15	13.2%	10.7%	16,6%
P50 (±50 million)	62.8%	P15-P85	72.4%	83.6%	75.6%
P85 (±50 million)	50.9%	≥ P85	28.9%	25.6%	30.6%
Sum	141.7%		114.5%	119.9%	122.4%

Table 2. *Costs that are appropriate in sentences featuring assorted verbal probability expressions (median values and percentage of participants suggesting extreme values (amounts less than minimum or more than assumed maximum values)*

	Study 1			Study 2: Group 1			Study 2: Group 2		
Expression	Mdn	≤ min	≥ max	Mdn	≤ min	≥ max	Mdn	≤ min	≥ max
More than	1000	**56.8%**	9.1%	935	**56.8%**	4.5%	800	**73.5%**	0.0%
Less than	1300	2.3%	**75.9%**	1300	6.8%	**68.2%**	1800	0.0%	**85.3%**
Probably (likely)	1100	0.0%	12.2%	1100	2.3%	9.3%	1100	2.9%	11.8%
Not probable	1200	**32.5%**	**45.0%**	1100	**34.9%**	**37.2%**	1550	**26.5%**	**58.8%**
Improbable	1000	**52.4%**	**33.3%**	1300	**23.3%**	**48.8%**	1125	**47.1%**	**44.1%**
Can (could) cost	1250	9.5%	**55.2%**	1200	6.8%	**36.4%**	1200	2.9%	**26.5%**
Not surprised	1300	2.4%	**59.5%**	1300	2.3%	**38.6%**	1350	0.0%	**51.2%**
Entirely possible	1275	2.4%	**50.0%**	1257	2.3%	**27.3%**	1225	0.0%	**26.5%**
Perhaps	1200	2.4%	28.6%	1200	2.3%	**27.3%**	1125	2.9%	14.7%

Table 3. Which costs are «entirely possible», and which expression is appropriate for describing the maximum value in an uncertainty distribution? Choice percentages from judges in two conditions. Study 3.

Statement	*Alternatives*	*Choice percentages*
Condition A (N = 196)		
It is entirely possible that it will cost ...	750 million (P15)	6.1%
	1100 million (P50)	14.3%
	1450 million (P85)	**79.6%**
Condition B (N = 160)		
It is ... that it will cost 1450 million.	a low probability	16.9%
	entirely possible	**68.1%**
	Probable	15.0%

Figure Captions

Figure 1. Graph illustrating a probability distribution, divided in segments corresponding to P15 (minimum), P50 (expected) and P85 (maximum), as presented to participants in all three studies.

Figure 2. Mean accuracy ratings (1-7) of cost estimates when actual costs are below, equal to or above estimated maximum values, all studies.

Figure 3. Verbal expressions preferred by journalist and in conversations with employer (percent of respondents in Study 1 and Study 2 combined).

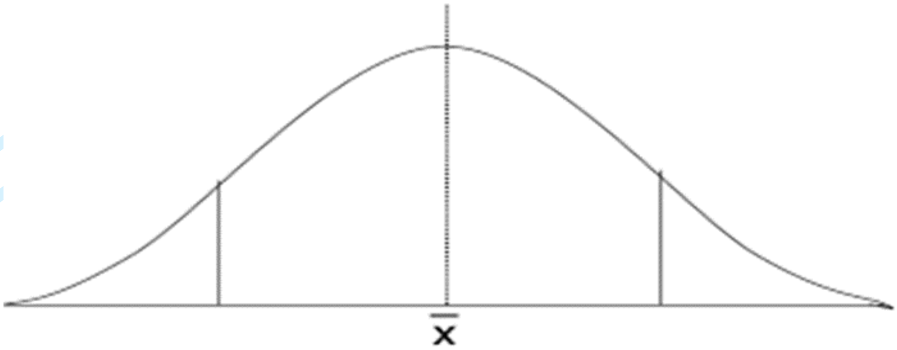

Figure 1. Graph illustrating a probability distribution, divided in segments corresponding to P15 (minimum), P50 (expected) and P85 (maximum), as presented to participants in all three studies.

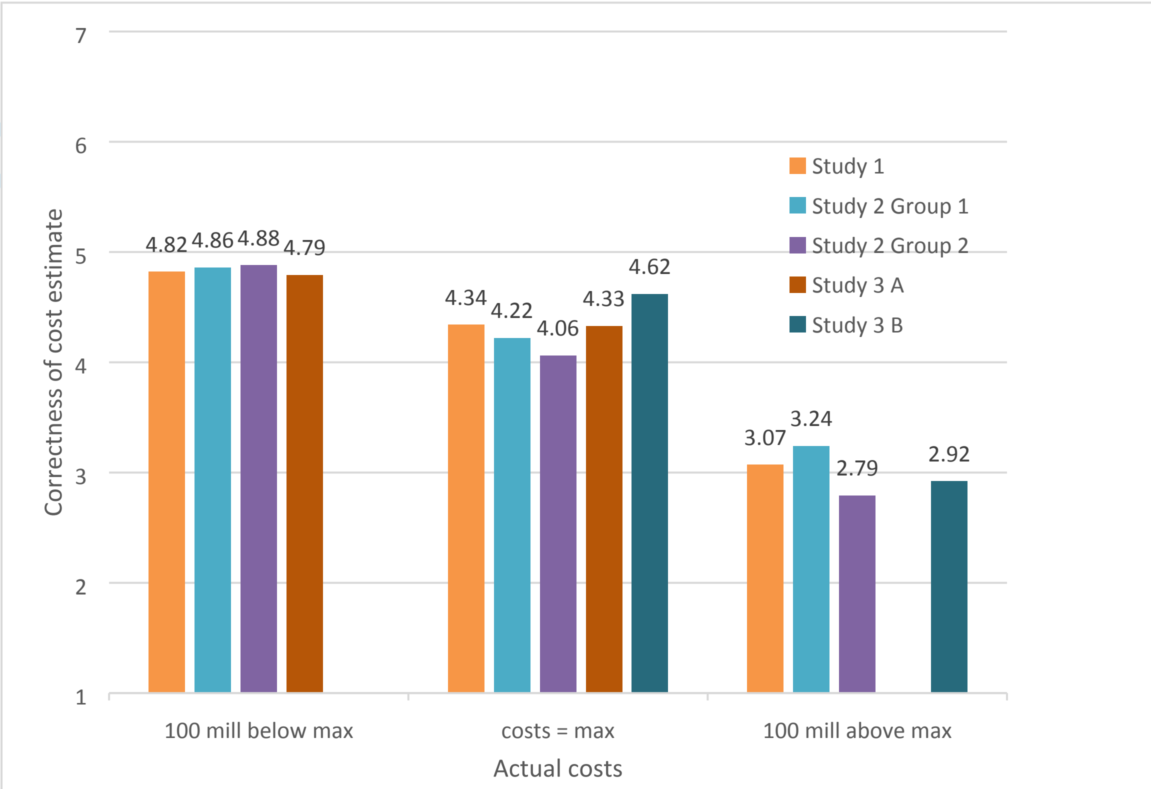

Figure 2. Mean accuracy ratings (1-7) of cost estimates when actual costs are below, equal to or above estimated maximum values, all studies.

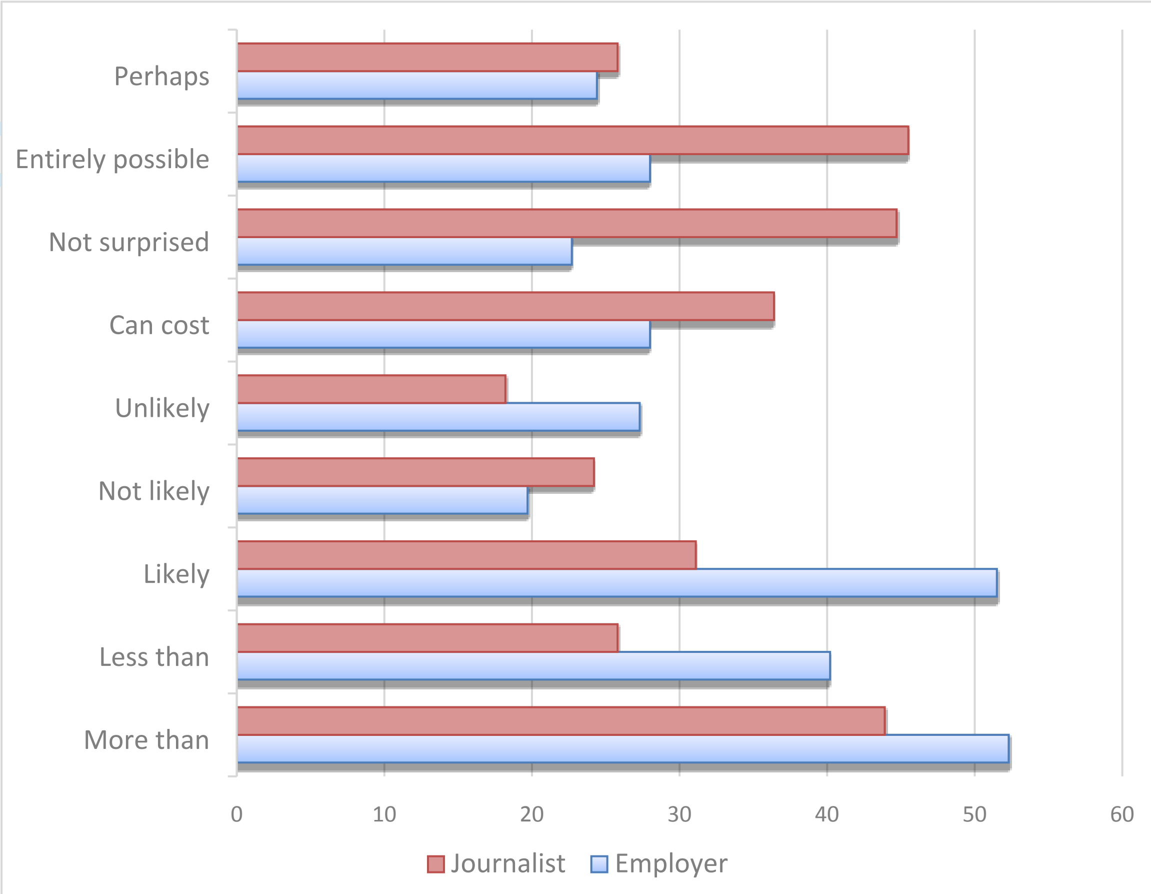

Figure 3. Verbal expressions preferred by journalist and in conversations with employer (percent of respondents in Study 1 and Study 2 combined).