Subjective probability assessment in decision analysis: Partition dependence and bias toward the ignorance prior

Craig R. Fox
The Anderson School of Management and Department of Psychology
University of California at Los Angeles

Robert T. Clemen
Fuqua School of Business
Duke University

Running Head: Partition Dependence of Subjective Probabilities

Address Correspondence to:
Robert T. Clemen
Fuqua School of Business
Duke University, Box 90120
Durham, NC 27708-0120
919-660-8005
clemen@duke.edu
Subjective probability assessment in decision analysis: Partition dependence and bias toward the ignorance prior

Abstract

Decision and risk analysts have considerable discretion in designing procedures for eliciting subjective probabilities. One of the most popular approaches is to specify a particular set of exclusive and exhaustive events for which the assessor provides such judgments. We show that assessed probabilities are biased toward a uniform distribution over all events into which the relevant sample space happens to be partitioned. This gives rise to judged probabilities that vary systematically with the partition of the sample space that is being evaluated. We surmise that a typical assessor begins with an “ignorance prior” distribution that assigns equal probabilities to all specified events, then insufficiently adjusts those probabilities to reflect his or her beliefs concerning how the likelihoods of the events differ. In five studies, we demonstrate partition dependence for both discrete events and continuous variables (Studies 1 and 2), show that the bias decreases with increased domain knowledge (Studies 3 & 4), and that top experts in decision analysis are susceptible to this bias (Study 5). We relate our work to previous research on the “pruning bias” in fault-tree assessment (e.g., Fischhoff, Slovic, & Lichtenstein, 1978) and show that previous explanations of pruning bias (enhanced availability of events that are explicitly specified, ambiguity in interpreting event categories, demand effects) cannot fully account for partition dependence. We conclude by discussing implications for decision analysis practice.

Key Words: Probability assessment, risk assessment, subjective probability bias, fault tree
1. Introduction

Decision and risk analysis models often require assessment of subjective probabilities for uncertain events, such as the failure of a dam or a rise in interest rates. Speztler and Staël Von Holstein (1975) were the first to describe practical procedures for eliciting subjective probabilities from experts. Their procedures are still in use, largely unchanged, as reflected in work by Clemen and Reilly (2001), Cooke (1991), Keeney and von Winterfeldt (1991), Merkhofer (1987), and Morgan and Henrion (1990).

Human limitations of memory and information processing capacity often lead to subjective probabilities that are poorly calibrated or internally inconsistent, even when assessed by experts (see, e.g., Kahneman, Slovic, & Tversky, 1982; Gilovich, Griffin, & Kahneman, 2002). In this paper we study a particular bias in probability assessment that arises from the initial structuring of the elicitation. At this stage the analyst, sometimes with the assistance of an expert, identifies relevant uncertainties and may partition each corresponding state space into a finite number of exclusive and exhaustive events for which probabilities will be judged. Although existing probability-assessment protocols provide guidance on important steps in the elicitation process (e.g., identifying and selecting experts, training experts in probability elicitation, the probability assessment itself), little attention has been given to the choice of specific events to be assessed.

In developing an elicitation structure, the analyst chooses the frame within which the expert assesses and communicates his or her probabilistic beliefs. If the uncertain event is defined by a continuous variable, the analyst may specify intervals for which the expert will assess probabilities. Sometimes these intervals are defined by salient reference points such as the status quo or target values. For instance, an expert might be asked to assess probabilities that next quarter’s sales will rise or fall relative to their current level or to assess the probabilities that completion time for a project will exceed or fall within the budgeted time. Intervals may also be dictated by thresholds relevant to the decision. For instance, a farmer might assess the probabilities that next year’s crop price will exceed or fail to exceed 38 cents per bushel because only a price above that threshold would justify the purchase of an adjacent plot of land. If no obvious threshold values exist, the analyst and expert must agree on a more arbitrary set
of intervals. For example, the expert may be asked to evaluate the probabilities that first year sales on a new product will fall in the following ranges: low (0 to 1000 units), medium (1001-2000 units), and high (more than 2000 units).

Analysts typically assume that the particular choice of intervals does not unduly influence assessed probabilities. Unfortunately, our experimental results demonstrate that this assumption is unfounded: assessed probabilities can vary substantially with the particular partition that the analyst chooses. We refer to this phenomenon as partition dependence (see also Fox & Rottenstreich, 2003). It is more general than the pruning bias documented in the assessment of fault trees by Fischhoff, Slovic, and Lichtenstein (1978) (hereafter FSL), in which particular causes of a system failure (e.g., reasons why a car might fail to start) are judged more likely when they are explicitly identified (e.g., dead battery, ignition system) than when pruned from the tree and relegated to a residual catch-all category (“all other problems”). Most previous investigators have interpreted pruning bias as an availability or salience effect: when particular causes are singled out and made explicit rather than included implicitly in a catch-all category, people are more likely to consider those causes in assessing probability; as FSL put it, “what is out of sight is also out of mind” (p.333).

Our goal in this paper is to extend the investigation of pruning bias from fault trees to the more general problem of probability assessment of event trees. Our studies suggest that the traditional availability-based account does not fully explain pruning bias or the more general phenomenon of partition dependence. We propose an alternative mechanism: a judge begins with equal probabilities for all events to be evaluated and then adjusts this uniform distribution based on his or her beliefs about how the likelihoods of the events differ. Bias arises because the adjustment is typically insufficient. Although current best practices in subjective probability elicitation are designed to guard against availability and the other major causes of pruning bias that have been previously advanced in the literature, such best practices provide inadequate protection against a more pervasive tendency to anchor on equal probabilities. Understanding the nature and causes of partition dependence can help analysts identify
conditions under which this bias may arise, predict conditions that may exacerbate or mitigate the effect, and develop more effective debiasing techniques.

In the following section of this paper we review literature on pruning bias and partition dependence. In §3 we describe a series of studies that document the robustness of partition dependence across a variety of contexts beyond fault trees, provide support for our interpretation of this phenomenon, and cast doubt on the necessity of alternative accounts that have been proposed to explain pruning bias. We close with a discussion of the interpretation and robustness of partition dependence, other manifestations of this phenomenon, and prescriptive implications of our results.

2. Literature Review

FSL presented professional automobile mechanics and laypeople with trees that identified several categories of reasons why a car might fail to start as well as a residual category of reasons labeled “all other problems.” Participants were asked to estimate the number of times out of 1000 that a car would fail to start for each of the categories of causes specified. When the experimenters removed (pruned) specific categories of causes from the tree (e.g., fuel system defective) and relegated them to the residual category as in Figure 1, the judged probability of the residual category, as assessed by a new group of participants, did not increase by a corresponding amount. Instead, the probability for the categories that were pruned from the tree tended to be distributed across all of the remaining categories. Because the probability assigned to the residual category in the pruned tree was lower than the sum of probabilities of corresponding events in the unpruned tree, the pattern has subsequently come to be known as the pruning bias (e.g., Russo & Kolzow, 1994).

Since the publication of FSL, numerous authors have replicated and extended the basic result and proposed three major explanations for pruning bias: availability, ambiguity, and credibility. Below we review each of these accounts.

Availability. In explaining pruning bias, FSL invoked the availability heuristic (Tversky & Kahneman, 1973): judged probabilities depend on the ease with which instances can be recalled or scenarios constructed. In the case of fault trees, explicitly mentioning a cause or category of causes will
make that cause or category more salient, easing retrieval of related instances or construction of relevant scenarios, and hence leading to an increase in the corresponding judged probability. Support for such a mechanism has been provided by a number of researchers since FSL, notably Van der Pligt, Eiser, and Speark (1987), Dubé-Rioux and Russo (1988), Russo and Kolzow (1994), and Ofir (2000).  

**Ambiguity.** Hirt and Castellan (1988) argued that some categories of problems in FSL’s automobile fault tree are ambiguous. For example, suppose that the branch labeled “battery charge insufficient” were removed from the tree. Specific causes that might fit into that category, such as “faulty ground connection” or “loose connection to alternator,” could just as well be assigned to a remaining branch labeled “ignition system defective” as to the residual “all other causes” category. Such ambiguous mapping of specific causes to categories could give rise to the observed pattern in which probabilities of pruned branches are distributed across remaining branches.

**Credibility.** A third explanation of the pruning bias is that people assume a credible real-world fault tree would list enough possible causes so that the catch-all category would be relatively unlikely, and each explicitly listed cause should have a nontrivial probability (Dubé-Riou & Russo, 1988; see also FSL, pp. 340-341). This argument suggests that the pruning bias represents a demand effect (Clark, 1985; Grice, 1975; Orne, 1962), whereby a participant considers the assessment as an implicit conversation with the experimenter in which the experimenter is expected to adhere to accepted conversational norms, including the expectation that any contribution should be relevant to the aims of the conversation. In the case of fault trees, the probability assessor may presume that any branch (other than the catch-all) for which a probability is solicited must have a nontrivial probability; otherwise the probability of that item would be irrelevant and therefore the query would violate conversational norms. FSL were able to cast

---

1 Ofir (2000) noted that the original characterization of the availability heuristic (Tversky & Kahneman, 1973) is that people sometimes judge likelihood by ease of retrieval (i.e., how readily instances come to mind) and not the content of retrieval (i.e., the number of instances retrieved; see Schwarz et al., 1991). His data suggest that people with less domain knowledge rely on the ease with which they can retrieve specific causes (i.e., the availability heuristic), whereas people with more domain knowledge are influenced by the absolute number of specific causes that come to mind. Regardless of how an expert assesses likelihood (by ease of retrieval, content of retrieval, or some other mechanism), the availability-based account of pruning bias holds that specific causes or events are more likely to be considered when they are explicitly identified than when they are implicit constituents of a superordinate category.
doubt on the credibility account in their studies, because the mean probability assigned to the least important of seven branches was only 0.033, and the catch-all category received a higher mean probability than the least probable identified category (Study 1).

In an attempt to disentangle the roles of availability, ambiguity, and credibility as potential explanations of pruning bias, Russo and Kolzow (1994) presented MBA students with two full or pruned fault trees that listed 7 or 4 categories of fault including a catch-all: one tree entailed causes of death for a randomly selected individual and the other was FSL’s reasons for a car not starting. Participants then completed three steps: (1) They generated more specific causes for each branch, (2) they classified a standard set of specific causes into categories, and (3) they estimated the likelihood of each branch. The authors argued that the availability explanation attributes bias to the first stage (hypothesis generation), the ambiguity explanation attributes bias to the second stage (hypothesis categorization), and the credibility explanation attributes bias primarily to the third stage (likelihood assessment). Results replicated the pruning bias for both trees and found a concomitant effect on the number of specific causes that could be generated (the more causes that came to mind the higher the judged probability), thereby supporting an availability-based interpretation. Although participants misclassified more than half of the causes in the automobile tree, they misclassified only one cause in seven for the causes-of-death tree, suggesting that ambiguity is not necessary to produce pruning bias. One group of participants was told that the categories were constructed by a panel of experts, a second group was told that they had been generated from responses of a group of 15-year-old students, and a third (control) group was told nothing about the source of the tree. The credibility account suggests that the catch-all category should receive relatively low probabilities in the control condition, roughly the same probabilities in the “expert” condition (because the credibility account assumes the default should be an expert tree), and somewhat higher probabilities in the “student” condition (because young students do not necessarily produce credibly complete trees). Instead, Russo & Kolzow (1994) found that the catch-all category was assigned similar probabilities in the student and control conditions and it was assigned somewhat lower probability
in the expert condition. Thus it appears that when the source of the tree was not identified the perceived credibility of the trees was not a significant factor underlying the observed pruning bias.

Although the three foregoing accounts (availability, ambiguity, credibility) may all contribute to some instances of pruning bias, and although availability appears to be the most robust mechanism, we assert that even availability does not provide an adequate explanation of pruning bias. In particular, the availability account predicts that there should be little or no effect of pruning causes from a full tree if these causes are explicitly mentioned as part of the catch-all category (so that the pruned causes are no longer out of sight even though their probabilities are not assessed separately). However, when FSL did this (Study 5) they nevertheless observed a strong pruning bias—a result that has received surprisingly little subsequent attention in the literature and which begs for a new interpretation of the phenomenon.

**Anchoring and insufficient adjustment.** We propose a fourth mechanism driving pruning bias: people anchor on a uniform distribution of probability across all branches of the fault tree and adjust according to features that distinguish each branch. Because such adjustment is usually insufficient (Tversky & Kahneman, 1974; Epley & Gilovich, 2001), assessors are biased toward probabilities of 1/n for each of n branches in the tree. To illustrate, consider a fault tree consisting of seven branches plus a residual category. According to the anchoring account, the assessed probability of the residual will be biased toward 1/8 because it is one branch of eight. Now imagine pruning this tree so that three branches remain, plus a residual category. Although the residual subsumes five of the original branches, it now represents a single branch of four. The anchoring account predicts that the assessed probability of the residual in this pruned tree will be biased toward 1/4 rather than 5/8 and that the remaining branches will be biased toward 1/4 rather than 1/8.

Starting with equal probabilities for all branches can be interpreted as an intuitive application of the so-called “principle of insufficient reason” that has been attributed to Leibniz and Laplace (Hacking, 1975). We say that a probability assessor adopts an ignorance prior, by which we mean a default judgment that category probabilities are equal. Taking equal probabilities as a starting point, a probability assessor then adjusts (usually insufficiently) to account for his or her beliefs about how the likelihood of
the events differ. Although we interpret our results in terms of anchoring and insufficient adjustment, a bias toward the ignorance prior may also be driven in some cases by enhanced accessibility of information that is consistent with an equal distribution of probability (Chapman & Johnson, 2002) or the intrusion of error variance into the processing of frequency information (Fielder & Armbruster, 1994).

The anchoring hypothesis has not been extensively investigated and the existing empirical evidence for it is rather indirect. Van Schie and van der Pligt (1990) asked undergraduates to estimate the proportion of acid rain that could be attributed to various causes and found that the cause “traffic” received a median rating of 14% in a (full) eight-branch tree and a median rating of 24% in a (pruned) four-branch tree, very close to the corresponding ignorance prior probabilities of 1/8 and 1/4, respectively. Johnson, Rennie, and Wells (1991) asked undergraduates to judge the relative frequency of possible outcomes when a baseball player is at bat (e.g., single, double, out), the true values of which were known to the experimenters. Participants tended to underestimate relative frequencies when the corresponding ignorance prior was below the true value and overestimate when the corresponding ignorance prior was above the true value. Harries and Harvey (2000, pp. 441-442) obtained a similar result using a causes-of-death probability estimation task. Russo and Kolzow (1994, p. 26, footnote 13) asked participants “what should be” the probability of a residual category for a typical tree with different numbers \( n \) of labeled branches and observed that responses provided a “remarkable fit” to the formula \( p_n = 1/(n+1) \), the ignorance prior.

In the section that follows we offer more direct evidence that pruning bias is driven by a tendency to allocate probability evenly across all events into which the event space happens to be partitioned. In five experiments we extend the observation of partition dependence from the narrow domain of fault trees (judgments of the relative frequency of various categories of fault in a system) to the more general domain of assessed probabilities of uncertain events. We demonstrate that even sophisticated probability assessors are susceptible to partition dependence in situations where the availability, ambiguity, and credibility mechanisms can be largely ruled out. Thus, we show that reliance on ignorance priors is the most robust source of partition dependence and that bias in subjective probability assessment may be more prevalent than has been previously supposed.
Of course the mechanisms four mechanisms reviewed here are not mutually exclusive. But to the extent that pruning bias is driven by the traditional explanations discussed in the literature (availability, ambiguity, credibility), existing best practices should mitigate their impact. For example, conditioning of experts (e.g., Merkhofer, 1987) draws out extensive knowledge about the topic at hand and hence may reduce availability effects. Use of the clarity test (Howard, 1988) is designed explicitly to banish ambiguity of categories from the assessment task. Involving the expert in structuring the elicitation may reduce any potential credibility effects. However, to the extent that pruning bias is driven by a more general tendency to anchor on the ignorance prior, none of these best practices will protect the procedure against systematic bias, and new corrective procedures will be called for.

3. Experimental Evidence

Study 1: Separate evaluation of events trumps separate description of events.

Most studies of fault trees have confounded whether particular causes were explicitly identified with whether participants were asked to assess probabilities of those causes. A straightforward reading of the availability account predicts that the probability assigned to a particular category will increase when it is explicitly identified in the tree but will not be affected by whether it is evaluated separately or with other causes. In contrast, the ignorance prior account predicts that the distribution of probabilities will be affected primarily by the number of branches that are explicitly evaluated. As mentioned earlier, some studies (including FSL’s Experiment 5) have found that, holding descriptions constant, events are generally assigned higher probabilities when split into multiple branches that are evaluated separately. Likewise, in their formulation of support theory, Rottenstreich and Tversky (1997) found that although unpacking a category (e.g., homicide) into a disjunction of subcategories (e.g., homicide by an acquaintance or homicide by a stranger) generally increases judged probability, separate assessment of the subcategories increases aggregate judged probabilities still further. A subsequent review of several studies (in Sloman, et al., 2004) found that the effect of separate evaluation is more robust and more pronounced than that of unpacking the description. This pattern is consistent the notion that judged probabilities are affected more by a bias toward 1/2 for each event that is evaluated (1/2 is the ignorance
prior when considering an event against its complement) than by the enhanced availability of constituent events when the description is unpacked.

Our first study was designed to demonstrate in the context of event trees that the effects of separate evaluation (predicted by the ignorance prior account) persist even when the effects of unpacking the description (predicted by the availability account) are negligible. Unlike previous fault-tree studies cited above, we asked participants to judge the probabilities of future events, and we used well-defined categories whose constituents were well known to participants, rendering the ambiguity account less relevant.

**Method.** We recruited 93 weekend MBA students at Duke University mid-way through a required course on decision models. By the time the study was run, participants had already learned about basic decision analysis tools including decision trees and subjective probability-assessment methods. All participants had previously completed an MBA course on probability and statistics.

Participants judged probabilities that particular schools would receive the top spot in *Business Week*’s next biennial ranking of business schools, a topic with which we expected them to be very familiar\(^2\). Each participant read the following instructions:

In the most recent *Business Week* rankings of daytime MBA programs, the Wharton School was ranked #1. In each of the spaces provided below, please write your best estimate of the probability that the daytime MBA program(s) indicated will be ranked #1 in the next *Business Week* survey... Please make sure that your probabilities sum to 100%.

Participants in the *full-tree* condition \((n = 30)\) were then presented with a tree in which the strongest MBA programs (plus a catch-all category) were listed alphabetically on separate branches:

- Chicago
- Harvard
- Kellogg
- Stanford
- Wharton
- None of the above

\(^2\) Fuqua administrators had previously conducted a survey of students admitted to Duke’s daytime MBA program \((N = 285)\), and 99% of respondents indicated that they had used *Business Week* and/or *US News and World Report*’s published rankings of business schools in deciding which business school to attend. Although our weekend MBA participants may have been somewhat less familiar with the details of the *Business Week* ranking than the daytime MBA students, we believe that our participants knew enough about this topic to make informed judgments in our study.
Participants in the collapsed-tree condition \((n = 32)\) were presented with a tree in which the residual category had been unpacked to remind participants of the same schools:

- Chicago, Harvard, Kellogg, Stanford, or another school other than Wharton
- Wharton

Participants in the pruned-tree condition \((n = 31)\) were presented a tree that included the following branches:

- A school other than Wharton
- Wharton

We predicted that unpacking the pruned tree into the collapsed tree would have minimal effect on participants’ judged probabilities of the residual category because we would be reminding experts of schools that should be salient to them even without explicit prompting. However, we predicted that expanding the collapsed tree into the full tree would substantially increase the aggregate judged probability of schools other than Wharton because the ignorance prior increases from 1/2 to 5/6.

**Results and discussion.** The results of Study 1 are displayed in Table 1 and accord with our predictions. The pruned and collapsed conditions both yielded median probabilities of 0.40 for the “other” (i.e., not Wharton) category. However, when asked to judge events separately in the full condition, the median sum of probabilities for schools other than Wharton jumps to 0.70. Based on a one-tailed Wilcoxon rank-sum statistic (which we use hereafter unless otherwise indicated), the median sum of judged probabilities for the full-tree is significantly different from median judged probabilities of the corresponding events in the collapsed and pruned conditions \((p = .05\) and \(p = .005\), respectively). Judged probabilities for a school other than Wharton in the collapsed and pruned conditions do not differ significantly \((p = .35)\).

The results for the school rankings replicate findings of FSL (Experiment 5) and Rottenstreich & Tversky (1997) that the judged probability of an event is higher when constituent events are assessed separately than when they are assessed as a single composite event. Furthermore, our results suggest that the availability-based account is not a necessary source of the pruning bias. In both the implicit (pruned)
and explicit (collapsed) disjunctions for which schools other than Wharton comprise one of two branches, median judged probabilities were slightly below the ignorance prior of 1/2. In the separate evaluation (full) condition, for which schools other than Wharton comprise five of six branches, the median sum of probabilities is slightly below the ignorance prior of 5/6.

**Study 2: Ignorance gives rise to strong partition dependence.**

Decision and risk analysts strive to find knowledgeable experts to provide probability assessments. Of course, analysts must often obtain assessments concerning unfamiliar or unprecedented future events, for instance in situations involving the development of a new technology or the management of an unproven hazard. The ignorance prior account suggests that partition dependence will be most pronounced in situations where probability assessors have little relevant knowledge and therefore have little basis to adjust probabilities from the ignorance prior. In our second study, we asked business students to make judgments and decisions concerning the future closing value of the Jakarta Stock Index (JSX), a domain about which we expected them to know very little. We reasoned that if we could observe partition dependence for the JSX, it would be difficult to attribute this bias to an availability-based mechanism for two reasons: (1) the extension of our categories (i.e. the set of possible closing values to which each range refers) – is readily apparent and therefore unpacking into subranges will only remind participants of subcategories that were patently obvious in the original tree; (2) participants cannot easily judge likelihood by availability of instances because it is unlikely that these participants can recall any instance of closing values of the JSX. Of course, one could argue that judged probabilities under ignorance are arbitrary and not a valid measure of respondents’ belief strength. In order to provide concomitant evidence that these judged probabilities accord with subjective degrees of belief, we also asked participants to make choices involving these events using an incentive-compatible payoff mechanism.

**Method.** Participants were 246 entering MBA students at Duke University who were asked during their orientation to complete a number of unrelated faculty research projects in exchange for a donation to a charity. All participants were presented with the following information:
The JSX is the leading composite index of the Jakarta Stock Exchange. The closing value of the JSX on December 31 of this year will be in one of the following ranges:

Approximately half the participants were then presented with the following ranges:

A) less than 500
B) at least 500 but less than 1000
C) at least 1000.

Participants in the *three-fold low* condition (*n* = 58) were asked to judge the probability the JSX would close in either range *A* or *B*. Participants in the *three-fold high* condition (*n* = 61) were asked to judge the probability that the JSX would close in range *C*. The remaining participants were instead presented with the following ranges that entailed a refined partition of values above 1000:

a) less than 500
b) at least 500 but less than 1000
c) at least 1000 but less than 2000
d) at least 2000 but less than 4000
e) at least 4000 but less than 8000
f) more than 8000.

Participants in the *six-fold low* condition (*n* = 65) were asked to judge the probability that the JSX would close in either range *a* or *b*. Participants in the *six-fold high* condition (*n* = 62) were asked to judge the probability that the JSX would close in range *c*, *d*, *e*, or *f*.

After providing a probability judgment, all participants were asked whether they would prefer to receive $10 for sure or receive $30 if the actual value of the JSX on the previous day had fallen into the specified interval (and receive nothing otherwise). We told participants that one respondent would be randomly selected to have his or her choice honored for real money.

**Results and discussion.** Figure 2 displays the results of Study 2. Judged probabilities varied dramatically by experimental condition, consistent with the ignorance prior account. The median judged probability that the Jakarta Stock Index (JSX) would close below 1,000 was 0.67 in the *three-fold low* condition (in which this event comprised two of the three specified ranges) but only 0.30 in the *six-fold low* condition (in which this event comprised two of six specified ranges), a significant difference (*p* = .02). Similarly, the median judged probability that the JSX would close at 1,000 or above was 0.25 in the *three-fold high* condition (in which this event comprised one of three specified ranges) and 0.60 in the *six-fold*
high condition (in which this event comprised four of six specified ranges), again a significant difference ($p = .001$). In three of four conditions judged probabilities did not differ significantly from the corresponding ignorance prior. Using binomial tests and distributing ties evenly, $p = .69$ in the 3-fold low condition, $p = .17$ in the 6-fold low condition, $p = .04$ in the 3-fold high condition, and $p = .25$ in the 6-fold high condition.

Results from the choice task echo the judged probabilities. A majority of participants in the three-fold low condition (55%) indicated that they would rather receive $30 if the JSX had closed below 1000 than receive $10 for sure, whereas a minority of participants in the six-fold low condition (31%) made the same choice ($\chi^2(1) = 7.48, p = .006$). Likewise, only 28% of three-fold high participants indicated that they would rather receive $30 if the JSX had closed at 1000 or above, whereas 58% in the six-fold high participants made the same choice ($\chi^2(1) = 11.43, p = .001$).

**Study 3: Domain knowledge moderates partition dependence.**

The first two studies establish that partition dependence can occur in situations where availability-based explanations are dubious at best. In the next study we examine the extent to which domain knowledge moderates this phenomenon. FSL, Ofir (2000), and Harries and Harvey (2000) showed in the context of fault trees that the pruning bias is reduced but not eliminated as domain knowledge increases. The ignorance prior account implies more generally that increasing knowledge should be associated with less reliance on the ignorance prior distribution (i.e., more adjustment) and hence probabilities that are less partition-dependent. We asked MBA students for probabilities relating to two domains for which we expected them to have very different levels of knowledge. One domain was the starting salary of graduates from their program, a topic that is closely followed by MBA students. The second domain was the starting salary of Harvard Law graduates, a topic that we expected to be much less familiar to our participants.

**Method.** The participants in this study were 120 second-year MBA students at Duke enrolled in an elective course in decision analysis. At the time of the study, these students had finished a first-year internship and were actively seeking permanent jobs. All participants had previously completed a course on probability and statistics and a course on decision models. The Duke MBA Career Services Office provides students with information about beginning salaries for graduates from previous classes.
The questionnaire asked participants to judge probabilities that the starting salary for a randomly chosen member of the current graduating class would fall into particular intervals. In order to construct roughly comparable partitions, we conducted a pretest in which a different sample of second-year MBA students assessed 10\textsuperscript{th}, 50\textsuperscript{th}, and 90\textsuperscript{th} percentiles for the first-year starting salary of a randomly selected member of the graduating class of Duke MBA students and the same for the present graduating class of Harvard Law students. Based on these assessments, we created low and high partitions for both Duke MBA and Harvard Law salaries that were roughly comparable. Participants in the low (high) partition condition provided probabilities for both Duke MBA and Harvard Law salaries, in which low (high) salary ranges were broken into sub-ranges, as displayed in Figure 3. In all cases we counterbalanced the order in which the two sets of probabilities were elicited. As before, participants were asked to ensure that their assessed probabilities for each variable summed to 100%. In addition to the probability judgments, we asked participants to rate their level of knowledge of the two variables on a scale from 0 (“I know nothing”) to 10 (“I know a great deal”).

**Results and discussion.** Median knowledge ratings were 7 for Duke MBA starting salaries (M = 6.78, SD = 1.87) and 2 for Harvard Law starting salaries (M = 2.03, SD = 2.00), confirming the validity of our a priori assumptions concerning relative knowledge. In fact, only one person of 120 indicated more knowledge about Harvard Law than Duke MBA; five others indicated the same degree of knowledge for both schools.

Table 2 presents results from Study 3. Before analyzing judged probabilities, we discarded responses from participants whose probabilities did not sum to 100%. The number of remaining responses for each cell is shown in Table 2. For a participant in the high partition, let $P_{\text{high}}(\text{Harvard}<90K)$ denote the single judged probability that a randomly chosen graduate of Harvard Law will earn less than $90,000 during his or her first year after graduation (the top entry in the lower right-hand cell in Table 2). Let $P_{\text{low}}(\text{Harvard}<90K)$ denote the corresponding sum of judged probabilities for a participant in the low partition (the sum of the top four entries in the upper right-hand cell in Table 2). Define $P_{\text{high}}(\text{Duke}<85K)$ and $P_{\text{low}}(\text{Duke}<85K)$ similarly. Median probabilities presented in Table 2 reveal partition dependence for
judgments of both Harvard and Duke salaries. In particular, judged probabilities are lower when they are derived from a single judgment than when they are derived from multiple judgments that are summed: $P_{\text{high}}(\text{Harvard<90K}) < P_{\text{low}}(\text{Harvard<90K})$ and $P_{\text{high}}(\text{Duke<85K}) < P_{\text{low}}(\text{Duke<85K})$. To perform an overall test for significance of the effect, we calculated $P_{\text{low}}(\text{Harvard<90K}) + P_{\text{low}}(\text{Duke<85K})$ for each participant in the low condition, and $P_{\text{high}}(\text{Harvard<90K}) + P_{\text{high}}(\text{Duke<85K})$ for each participant in the high condition. The ignorance prior account predicts that the median sum of probabilities below the relevant cutoff ($<85K, < 90K$) will be greater for participants in the low partition condition than the corresponding probability for participants in the high partition condition; this prediction is confirmed ($p < .0001$).

Thus, we observed substantial partition dependence for both schools, and this pattern was more pronounced for Harvard (difference of medians = 0.45) than for Duke (difference of medians = 0.35). To test the statistical significance of the interaction, we calculated $P_{\text{low}}(\text{Harvard<90K}) - P_{\text{low}}(\text{Duke<85K})$ among participants in the low partition conditions (for whom these values were the sums of four separate judgments) and $P_{\text{high}}(\text{Harvard<90K}) - P_{\text{high}}(\text{Duke<85K})$ among the participants in the high partition condition (for whom these values each refer to a single judgment). If partition dependence is more pronounced for the Harvard Law judgments than for the Duke MBA judgments, we would expect the difference to be larger for the low-partition respondents than for the high-partition respondents. This difference of differences approaches significance by a one-tailed Wilcoxon test ($p = .12$) and by a t-test ($t(105) = 1.63, p = .05$).

**Study 4: Credibility and demand effects do not adequately explain partition dependence.**

In Studies 1-3 we have shown that partition dependence is observed even in situations where availability and ambiguity accounts are unlikely to play a role, and that the effect is especially pronounced for less familiar domains in which one might expect greater reliance on the ignorance prior distribution. As mentioned earlier, one might be tempted to dismiss partition dependence, like the pruning bias, as a demand effect (Clark, 1985; Grice, 1975; Orne, 1962) or “credibility” effect (Dubé-Rioux & Russo, 1988). According to this interpretation, probability assessors derive information from the tree they
have been presented by assuming that it is a credible event tree in which each (non-residual) branch has a nontrivial probability of occurrence. Hence one might worry that the pruning bias and our demonstrations of partition dependence could be explained as an experimental artifact that would not appear if participants knew that the particular partitions they saw were arbitrary. In Studies 1-3 we have attempted to minimize these effects by using familiar domains (MBA program rankings Study 1, MBA salaries in Study 2) and financial incentives (Study 2). To address this issue more directly, we designed a new study in which participants could see clearly that they had been randomly assigned to a particular partition condition. An observation of partition dependence in this study would suggest that the credibility/demand-effect explanation cannot fully account for the phenomenon.

**Method.** We recruited 102 students enrolled in a decision models course in Duke’s Weekend Executive MBA program. Four participants were selected at random to receive $20 as a reward for completing a brief survey. Figure 4 displays the stimuli presented in the questionnaire. All participants were presented with two different four-interval partitions: the *low* partition had three intervals up to 4000 and one interval above 4000, and the *high* partition had one interval up to 4000 and three intervals above 4000.

Each participant assigned himself or herself to an experimental condition based on the last digit of his or her local telephone number. If the number was even (odd), the participant was asked to write “NASDAQ” above the left (right) tree and “JSX” above the right (left) tree. The order of presentation of *low* and *high* partitions was counterbalanced. Ultimately, \( n = 53 \) participants assigned themselves to the NASDAQ *low*, JSX *high* partition condition, and \( n = 49 \) participants assigned themselves to the NASDAQ *high*, JSX *low* partition condition. Below these trees we instructed participants to assign probabilities that the designated index would close in the specified range on the last day of trading of the present year. Participants were asked to ensure that their probabilities for each index summed to 100%. Finally, participants were asked to indicate their familiarity with each index on a 0 to 10 scale.

It should have been apparent to participants that they had assigned themselves at random to partition conditions. To the extent that participants read anything into the particular pair of partitions they
saw, they might have noticed the common 4000 threshold (at the time of the study the NASDAQ index was closing near 4000). Under the credibility account, noticing the common threshold might have biased probabilities above and below 4000 towards 50%. However this would also drive responses toward consistency across conditions and away from partition dependence. Thus, our design provides a conservative test of partition dependence in a situation where credibility and demand effects are not likely to play a role.

**Results and discussion.** Table 3 summarizes the results of Study 4. Median knowledge ratings were 7 for NASDAQ (M=6.34, SD = 2.67) and 0 for the JSX (M = 0.18, SD = 0.77), confirming our *a priori* expectation that the NASDAQ would be much more familiar than the JSX. Of 92 participants who provided knowledge ratings, only one rated the same level of knowledge for both NASDAQ and JSX (both zero). All others reported higher knowledge for NASDAQ.

Before analyzing the data we discarded responses that did not sum to 100% for each event tree. The number of responses remaining in each cell is shown in Table 3. Consider first the results for the full data set shown in the top section of Table 3. The overall pattern of partition dependence is highly significant (*p* < .0001), and it is significant for both the NASDAQ (*p* = .02) and JSX (*p* = .003) taken separately. The results for JSX closely replicate the results from Study 2. In fact, the median probability for all four intervals for JSX was 25% in both low and high partition conditions. More strikingly, modal judged probabilities under ignorance coincided precisely with the ignorance prior: 42% of respondents (41 of 97) provided equal probabilities for the four JSX intervals, and these responses were distributed roughly evenly across partition conditions. In light of the obvious random assignment of participants to experimental condition, these results indicate that partition dependence cannot easily be dismissed as a demand effect.

As in Study 3, rated knowledge seemed to moderate partition dependence. The difference between probabilities in high and low partition conditions was less pronounced for NASDAQ (difference in medians = 0.25) than for JSX (difference in medians = 0.50), and this interaction is highly significant (*p* = .002). To explore this knowledge effect further, we split the sample based on participants’ NASDAQ
knowledge ratings. Recall that the median rating was 7. Table 3 shows median judged probabilities among the “experts” with knowledge ratings of 7 or higher and “non-experts” with knowledge ratings below 7. Partition dependence is extremely pronounced among the non-experts (difference in medians = 0.68, \( p = .002 \)) but disappears among the experts, (difference in medians = -0.07, \( p = .44 \)), and the interaction is highly significant (\( p = .001 \)). There is no difference, however, between NASDAQ experts and non-experts on JSX assessments. These results provide further support for the claim that increased knowledge leads to a decrease in reliance on the ignorance prior. Although the self-rated experts concerning the NASDAQ did not fall prey to partition dependence in this instance, we suspect that this result is the exception rather than the rule, given the observations of partition dependence in other studies of participants with considerable substantive knowledge (e.g., Study 1, Study 3, the auto mechanics of FSL and Ofir, 2000). We also suspect that the juxtaposition of probability trees in Study 3 may have artificially cued some more savvy participants to take pains to be consistent in their use of probabilities across the two trees with which they happened to be presented.

**Study 5: Professional decision analysts are not immune.**

The results of Studies 1-4 demonstrate the robustness of partition dependence among a fairly sophisticated population: graduate students of business, most of whom had training in probability, statistics, and decision models (Studies 1, 3 and 4), and some of whom had additional training in decision analysis (Study 3). Despite this procedural sophistication, one could argue that the participants in our studies did not have extensive training in and experience using probability-assessment methods, and that such training and experience might eliminate the bias. In our final study, we address this issue by replicating the method of Study 4 using a population with extensive training in decision analysis: members of the Decision Analysis Society (DAS) of INFORMS.

---

3 To test the significance of the interaction, we considered the event NASDAQ < 4000. For this event, we pooled responses from low-partition experts and high-partition nonexperts, and likewise we pooled responses from high-partition experts and low-partition nonexperts. An interaction implies unequal medians for these two pooled groups. We used the Kruskal-Wallis statistic to test for differences in the medians of the two groups.
**Method.** We solicited responses from 169 members of the DAS email list, obtaining 58 responses. A portion of the questionnaire is shown in Figure 5. We asked respondents to judge probabilities concerning the membership totals of DAS and the Society of Quantitative Analysts (SQA) five years in the future. We told participants that the Society of Quantitative Analysts is “… a professional society incorporated in 1989 that is concerned with the application of innovative quantitative techniques in finance, investment and risk management.” At the time of the survey the DAS tallied 764 registered members, and the SQA reported on its website “over 200 members,” though we did not mention either of these facts in the instructions.

The first part of the survey followed exactly the same design as Study 4 in which we asked participants to assign themselves to conditions (low and high partitions) using the last digit of their primary home telephone number. The order of presentation of low and high partitions was counterbalanced. We obtained $n = 30$ participants in the DAS low, SQA high partition condition and $n = 28$ participants in the DAS high, SQA low partition condition. Participants then assessed the probabilities that the total membership of the DAS and SQA would fall into designated ranges. As usual, we asked participants to verify that their probabilities summed to one for each tree, and we counterbalanced the order of the trees. Following the probability assessments, we asked each participant for his or her highest level of education completed (e.g., BA, MS, ABD, PhD); whether he or she had taught a course in decision analysis; the number of applied decision or risk-analysis projects in which he or she had elicited probabilities over the previous two years and (if more than zero) what elicitation techniques were used.

**Results and discussion.** Our respondents collectively represent considerable decision analysis expertise. Of 57 usable responses, 86% had PhDs, 75% had taught at least one course in decision analysis, and 63% had elicited probabilities in a total of 156 applied DA projects in the previous two years.

Table 4 summarizes the results of Study 5. Note that some participants chose not to provide all of the probabilities requested; the number of usable responses for each cell is shown in the Table. The responses showed significant partition dependence overall ($p < .0001$), and this pattern is significant for judgments of DAS (difference in medians = 0.25, $p = .01$). The effect size was similar for SQA
(difference in medians = 0.25). Because responses for the latter society exhibited a great deal of noise, the pattern did not achieve statistical significance by a nonparametric Wilcoxon rank-sum test ($p = .45$), though it was significant by a parametric t-test ($t(44) = -2.28, p = .01$).

To explore the robustness of our results we examined a subsample of the 25 decision analysts with Ph.D. degrees who had worked on at least one applied decision analysis project in the past two years and had also taught at least one decision analysis course. Table 5 displays the results of this analysis. Not surprisingly, the overall effect of partition dependence is somewhat smaller among these super-experts, but the effect nevertheless approaches statistical significance (the difference in medians was 0.10 for DAS and 0.22 for SQA, overall $p = .05$ by one-tailed Wilcoxon rank-sum test).

Our main purpose in Study 5 was to demonstrate the robustness of partition dependence among a sample of participants with high procedural expertise. Although one might expect DAS members to know more about the size of DAS than SQA, our results here reveal no significant knowledge effect ($p = .30$) though there was a nonsignificant tendency among the super-experts. We did not collect knowledge ratings regarding the two societies, but we speculate that the lack of a significant knowledge effect may stem from small differences in knowledge across domains: We provided some information concerning the SQA, gave no information about current membership for either society, and asked about membership of both societies five years in the future.

4. General Discussion

In this paper we have extended the analysis of pruning bias from fault trees to the more general phenomenon of partition dependence in assessing subjective probability. In five studies using well-defined event trees we have accumulated support for the notion that this phenomenon is driven primarily by a bias toward equal allocation of probability across all events into which the sample space is partitioned, rather than the enhanced availability of events that happen to be made explicit, ambiguity of event categories, or information unintentionally conveyed by the particular branches that are selected for evaluation. In Study 1 we showed that unpacking the description of an event (a school other than Wharton will be the next top rated business school) into its most obvious constituents (Chicago, Harvard, Kellogg,
Stanford, or another school) did not lead to an increase in judged probability; however, asking participants to assess constituents separately gave rise to a dramatic increase in aggregate probability. In Study 2 participants displayed a pronounced degree of partition dependence in a situation where they were unlikely to have much knowledge (future close of the Jakarta Stock Index), with judged probabilities very close to corresponding ignorance prior probabilities, a tendency that was also reflected in betting behavior. Study 3 demonstrated that partition dependence among Duke MBA students was more pronounced for a relatively unfamiliar domain (Harvard Law graduate salaries) than a relatively familiar domain (Duke MBA graduate salaries). Study 4 replicated and strengthened the major findings of Studies 2 and 3 and also cast doubt on the credibility hypothesis that partition dependence is driven by a tendency of respondents to expect that all explicitly identified events have nontrivial probabilities. Finally, Study 5 demonstrated partition dependence among participants with considerable procedural expertise: members of the Decision Analysis Society. We close with a discussion of the interpretation and robustness of partition dependence, other manifestations of partition dependence, and prescriptive implications for decision and risk analysis.

The interpretation and robustness of partition dependence

Partition dependence refers to the tendency for judged probabilities to vary systematically with the way in which an event space is partitioned. Partition dependence thus appears to be analogous to framing effects in studies of choice (Kahneman & Tversky, 1984; Tversky & Kahneman, 1986) in which decisions are influenced by the way alternatives are described (e.g., in terms of losses and gains relative to a reference point). As with framing effects, respondents seem to accept the partition that is suggested to them in the form of an event tree, and they seem to be somewhat insensitive to the arbitrary nature of this partition. Our interpretation of this phenomenon is that people anchor their judgments on equal probabilities for each event in the specified partition (the ignorance prior distribution) and adjust insufficiently to account for their beliefs about how the likelihood of the events differ. We surmise that FSL’s major results concerning the impact of splitting and fusing branches of fault trees are driven by a
tendency to anchor on the ignorance prior, although we acknowledge that availability effects may also contribute.

Consistent with our anchoring-and-adjustment account, participants with greater substantive expertise show less partition dependence, and the effect may sometimes disappear when participants are particularly knowledgeable, especially if they are induced to consider multiple partitions of the event space (Study 4). This said, we believe that in many contexts experts may lack sufficient knowledge to overcome the bias. For instance, we believe that the MBA students in Study 3 were more knowledgeable about their future salaries than any other population would have been without explicit statistics at hand, and the knowledge ratings of these de facto experts confirmed a subjective feeling of high expertise. Perhaps more striking, partition dependence seems to be quite robust to varying levels of procedural expertise. It is difficult to imagine a population with greater knowledge of subjective probability assessment techniques than the DAS members surveyed in Study 5, yet even the most expert among them fell prey to partition dependence.

Other manifestations of partition dependence

Partition dependence has been observed not only in the context of event trees (in which assessors judge the probabilities of a number of exclusive and exhaustive events) but also in simple probability judgment. For example, Fox and Rottenstreich (2003) demonstrated that the language of a probability query can facilitate either a two-fold “case” partition {the target event obtains, the target event fails to obtain} and a corresponding ignorance prior of 1/2 or an \( n \)-fold “class” partition {event 1 obtains, event 2 obtains, … event \( n \) obtains} and a corresponding ignorance prior of 1/\( n \). For example, participants who were asked to judge the probability that “The temperature on Sunday will be higher than every other day next week” gave responses that tended toward 1/2, whereas participants who were asked to judge the probability that “Next week, the highest temperature of the week will occur on Sunday” gave responses that tended toward 1/7. In another set of studies See, Fox, and Rottenstreich (2004) demonstrated partition dependence in a learning environment where participants observed colored shapes that flashed on a computer screen with varying relative frequencies. When participants were then asked to judge the probability of a particular attribute (e.g., the probability of a black object versus the probability of a
triangle), they were biased toward the ignorance prior probability defined by the number of possible values that the target attribute could take (black was one of two possible colors, while a triangle was one of four possible shapes) even when these attributes appeared with identical objective frequencies. This bias diminished but did not disappear when participants had more extensive opportunities to learn the distribution of objects. Finally, Fox and Levav (2004) showed that common mistakes solving conditional probability puzzles such as the “Monty Hall” problem may reflect naïve extensional reasoning in which people subjectively partition the sample space on the basis of initial conditions, edit the partition using conditioning information, and calculate probability as a ratio of remaining events in the partition. They show further that subtle rewording of these problems can facilitate the use of more appropriate partitions and a higher frequency of correct responses. For instance, the question “Mr. Smith has two children, at least one of whom is a boy. What is the probability that the other child is a boy?” led most participants to invoke a naïve two-fold partition \{other child boy, other child girl\} and an incorrect response of 1/2, with very few participants providing the correct response of 1/3. However, the normatively equivalent question “Mr. Smith has two children and it is not the case that both are girls. What is the probability that both are boys?” led many more participants to invoke the more refined partition that identifies both children by birth order, \{boy-boy, boy-girl, girl-boy, girl-girl\}, edit out the last element, and provide the correct response of 1/3.

Partition dependence has been observed not only in likelihood judgment but also in other domains relevant to decision analysis. Weber, Eisenfuhr, and von Winterfeldt (1988) reported that when people are asked to assign weights to different attributes of potential outcomes, they assign greater overall weight if an attribute is split into component parts and weights are assessed separately for each component. This is consistent with a tendency to spread out weight relatively evenly among the attributes that happen to be identified. Benartzi and Thaler (2001) showed that when people make allocations of money among retirement savings instruments, they tend to diversify naively among the available options. For example, people offered a stock fund and a bond fund typically allocate half of their savings to each fund, while people offered a stock fund and a mixed stock/bond fund also typically allocate half of their savings to
each fund. Langer and Fox (2004) extended these results to allocation among simple chance lotteries and also find that allocations are influenced by the hierarchical organization of options (e.g., grouping of investments by vendor), which apparently influenced how the set was subjectively partitioned. Fox, Ratner, and Lieb (2005) extended the observation of partition dependence to riskless allocation of resources to beneficiaries (financial aid recipients, charities) and consumption opportunities to time periods. For a review of various manifestations of partition dependence in decision analysis, managerial resource allocation, and consumer choice, see Fox, Bardolet, and Lieb (in press).

The foregoing examples of partition dependence all involve the allocation of some scarce resource (probability, attribute weight, money) over a fixed set of possibilities (events, attributes, investments). All display a bias toward even allocation of the resource across the specified possibilities. These should be distinguished from superficially similar cases in which the weight assigned to an event depends on its rank relative to other events in the partition. For example, Birnbaum explains some event-splitting effects in risky choice (Starmer and Sugden, 1993) with his “transfer of attention exchange” (TAX) model, in which decision weight is transferred from higher-valued to lower-valued outcomes (see, e.g., Birnbaum, 2004). Similarly, Windschitl and Wells (1998) report that subjective likelihood of a focal event increases when the most likely alternative outcome is split into several less likely constituents.

**Prescriptive implications.**

In our survey of DAS members for Study 5, we asked what techniques these experts used in applied probability-elicitation projects involving continuous variables. Respondents reported that 58% of the time they rely on assessments based on pre-specified intervals (where intervals are either provided by the analyst or suggested by the expert) such as those used in this study. Another predominant technique is to elicit fractiles for the uncertain variable (often 10th, 50th, and 90th percentiles), a method that typically yields an overconfidence bias (e.g., confidence intervals for which the true value of the variable in question lies below the 10th percentile or above the 90th percentile more than 20% of the time; see Lichtenstein, Fischhoff, & Phillips, 1982; Klayman, Soll, Gonzalez-Vallejo, & Barlas, 1999). Further research is needed to determine whether some manifestation of partition dependence is observed in
fractile elicitation. But it is clear that both of these elicitation methods are susceptible to strong and persistent biases.

The present work has significant implications for improving existing best practices for eliciting subjective probabilities. Building on Russo & Kolzow’s (1994) process account, we suggest that assigning probabilities to event trees entails three subtasks, each of which may be susceptible to a distinct form of bias. First, experts must interpret the extension of each event to be evaluated—to what kinds of events does each branch refer? This stage may entail both the generation of possible constituent events and categorization of constituent events to branches of the tree. For instance in assessing the probabilities that a randomly selected death is due “disease,” “accident,” “homicide,” and “suicide,” an expert might need to subjectively elaborate the category “disease” by noting that it includes heart attacks, cancer, strokes, and various other diseases (after all one seldom learns that a person has died due to a generic “disease” and therefore may require more specificity to retrieve instances of such deaths from memory). Second, experts must evaluate support for each elementary event using judgmental heuristics, explicit arguments, computational models, historical frequencies, or some other approach. For instance a driver may assess the relative likelihood of various kinds of car failure by how easily instances of each category come to mind. Third, experts must map this impression of relative support into a set of numbers that sum to one. Such a decomposition may suggest specific corrective procedures that target biases attributed to each subtask.

Biases located at the first stage (interpretation of categories) may be driven by the availability and ambiguity mechanisms. Thus, one would expect these biases to be more significant for event spaces partitioned by category (e.g., different causes of death) than for event spaces partitioned along a single dimension (e.g., closing stock values) for which the interpretation of categories is transparent. For categorical trees the analyst can minimize this form of bias by working closely with the expert to carefully define and elaborate the interpretation of each branch.

Biases located at the second stage (assessment of support for each branch) can arise from a variety of sources that vary with the reasoning invoked by the expert. The analyst may be able to guard
against such biases to some extent by inducing the expert to articulate his or her reasoning, assumptions, and sources of information. Also, by working with the expert to develop an appropriate partition of the event space the expert should be less inclined to falsely infer that the analyst believes each event to have roughly equal support (i.e., this will minimize any potential credibility effects).

Biases located at the third stage (mapping assessed support onto a set of numbers) may be the most resistant to correction because they are least amenable to conscious reflection (cf. Arkes, 1991; Larrick, 2004). Because expert probability assessments can depend strongly on the specific partition used, the analyst should at the very least strive to direct the expert’s attention across the event space in as evenhanded a manner as possible. For categorical partitions, it will often be possible to tell whether one partition is more evenhanded than another. For instance, in judging the probability that one’s firm will win a competitive bid against a large number of competing firms, it may be convenient to assess the probabilities that (1) one’s own firm will win and (2) any one of the competing firms will win. However, we suspect that most people would consider a partition in which each competing firm’s chances are evaluated separately to be more evenhanded.

For continuous variables, it may be more difficult to determine what is an evenhanded partition. For instance, consider the decision of whether to launch a satellite on a particular day. Success may depend on the ambient temperature $T$, and it may be convenient to ask experts to assess probabilities that $T$ will fall above or below a specific target value (e.g., $T \leq 0^\circ C$ versus $T > 0^\circ C$). However, asking experts to assess probabilities that $T$ will fall in various specified intervals (e.g., $T \leq 0^\circ C$, $0^\circ C < T \leq 5^\circ C$, $5^\circ C < T \leq 10^\circ C$, $T > 10^\circ C$) may lead them to consider a more complete range of possible temperatures. Unfortunately, there may be little consensus concerning which set of intervals are the most evenhanded for most continuous variables.

Probability-elicitation procedures used in decision and risk analysis (Clemen & Reilly, 2001; Keeney & von Winterfeldt, 1991, Morgan & Henrion, 1990; Spetzler & Staël Von Holstein, 1975; von Winterfeldt & Edwards, 1986) can be thought of as instructions and devices to encourage deliberate and conscious reasoning. We paraphrase such best practice (described in detail in the references above) as,
“Elicit probabilities in a variety of ways and ask the expert to reconcile the inevitable inconsistencies among his or her judgments.” In particular, Spetzler and Staël Von Holstein (1975) describe several different approaches for assessing probability distributions for continuous variables, including fixing a value and asking for a cumulative (or exceedance) probability at that value; specifying a probability and asking for the corresponding fractile; asking for range estimates (e.g., 10\textsuperscript{th} and 90\textsuperscript{th} percentiles); and using the interval-splitting method. They show how the results from such a set of questions can lead to inconsistent probabilities, indicating the need to have the expert reconcile these differences. Responses from our survey of decision analysis experts in Study 5 are encouraging in this respect: nearly half (49%) of these experts indicated that they sometimes use multiple methods for eliciting subjective probabilities, and 12% indicated that they always use at least two methods.

Our results suggest an extension of Spetzler and Staël Von Holstein’s (1975) advice to include the use of multiple partitions as well as multiple assessment methods. Using multiple partitions can highlight inconsistencies that may arise due to reliance on different ignorance priors, and the analyst can then help the expert recognize and reconcile those inconsistencies. The particular partitions that should be used will naturally vary case-by-case. Thus, it may make sense for the analyst and expert to work together to specify several different partitions and explicitly compare results.

We note, however, that consistency or coherence is only one criterion by which to evaluate the quality of an expert’s subjective probabilities. Others include calibration and resolution (see, for example, Yates, 1990). Scoring rules are often used to evaluate probabilities, and the Brier score (Brier, 1950) in particular can be broken down into calibration and resolution components. Further research is needed to determine whether using multiple partitions can lead to improvements in these scores. Although we would expect to see some improvement due simply to the reconciliation of inconsistencies, whether the use of multiple partitions leads to improvements in calibration and resolution is an open question.

An alternative approach is to debias the judgments rather than the judge; doing so requires an explicit model of anchoring on the ignorance prior. Fox & Rottenstreich (2003), for example, introduce a model that is a refinement of support theory (Tversky & Koehler, 1994) in which judged probabilities are
a compromise between the balance of support for the target hypothesis (relative to its complement) and
the ignorance prior. One could in principle estimate the parameters of such a model and back out
subjective probabilities that are untainted by reliance on the ignorance prior. Further empirical work is
necessary to determine the adequacy of the model and the feasibility of this approach in decision analysis
practice.

Acknowledgements

This research was supported in part by NSF grants SES-00-99209 to Fox and SES-00-84383 and SES-03-
17867 to Clemen. Most of the work reported here was completed while Fox was on the faculty of the
Fuqua School of Business at Duke University, whose support is gratefully acknowledged. The authors
thank David Bardolet, Yuval Rottenstreich, Kelly See, and Canan Ulu for helpful discussions and
feedback.

References


Birnbaum, M. H., & Martin, T. (2003). Generalization across people, procedures, and predictions:
Violations of stochastic dominance and coalescing. In S. Schneider & J. Shanteau (Eds.), Emerging
Perspectives on Judgment and Decision Research. Cambridge, UK: Cambridge University Press.

1-3.

Chapman, G.B. & Johnson, E.J. (2002). Incorporating the irrelevant: Anchors in judgments of belief and
value. Chapter 6 in T. Gilovich, D. Griffin, & D. Kahneman (Eds.), Heuristics and biases: The
psychology of intuitive judgment. Cambridge, UK: Cambridge University Press.


Table 1. Study 1 Results

<table>
<thead>
<tr>
<th>Condition</th>
<th>Median $P$(School other than Wharton is ranked #1)</th>
<th>Significance levels using Wilcoxon rank-sum test (one-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Comparing</td>
</tr>
<tr>
<td>Pruned</td>
<td>0.40</td>
<td>$Pruned$ vs. $Collapsed$</td>
</tr>
<tr>
<td>Collapsed</td>
<td>0.40</td>
<td>$Pruned$ vs. $Full Tree$</td>
</tr>
<tr>
<td>Full tree</td>
<td>0.70</td>
<td>$Collapsed$ vs. $Full Tree$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$p$-value</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.350</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.053</td>
</tr>
</tbody>
</table>

Table 2. Median Knowledge Ratings and Probabilities from Study 3. The entries shown in bold italics median sums of separate probability assessments for four intervals. Columns labeled “$n$” indicate the number of usable responses for the corresponding cells.

<table>
<thead>
<tr>
<th>Knowledge Rating</th>
<th>Duke MBA</th>
<th>Harvard Law</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7/10</td>
<td>2/10</td>
</tr>
<tr>
<td>$\leq$ $85,000$</td>
<td>.75</td>
<td>.75</td>
</tr>
<tr>
<td>&gt; $85,000$</td>
<td>.25</td>
<td>.25</td>
</tr>
<tr>
<td>$n$</td>
<td>55</td>
<td>57</td>
</tr>
<tr>
<td>$\leq$ $90,000$</td>
<td>.75</td>
<td>.70</td>
</tr>
<tr>
<td>&gt; $90,000$</td>
<td>.25</td>
<td>.25</td>
</tr>
<tr>
<td>$n$</td>
<td>57</td>
<td>58</td>
</tr>
</tbody>
</table>

Table 3. Median Knowledge Ratings and Probabilities from Study 4. “Low partition” refers to the partition with intervals below 1000, 1001-2000, 2001-4000, and above 4000. “High partition” refers to the partition with intervals 4000 and below, 4001-8000, 8001-16000, and above 16000. Thus, the median probabilities shown in bold italics result from combining separate probability assessments for three intervals. Columns labeled “$n$” indicate the number of usable responses for the corresponding cells.

<table>
<thead>
<tr>
<th>Knowledge rating</th>
<th>NASDAQ</th>
<th>JSX</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7/10</td>
<td>0/10</td>
</tr>
<tr>
<td>$\leq$ 4000</td>
<td>.50</td>
<td>.25</td>
</tr>
<tr>
<td>&gt; 4000</td>
<td>.50</td>
<td>.25</td>
</tr>
<tr>
<td>$n$</td>
<td>49</td>
<td>51</td>
</tr>
<tr>
<td>Full data set</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low partition</td>
<td>0.75</td>
<td>0.75</td>
</tr>
<tr>
<td>High partition</td>
<td>.25</td>
<td>.75</td>
</tr>
<tr>
<td></td>
<td>51</td>
<td>51</td>
</tr>
<tr>
<td>NASDAQ “experts”</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low partition</td>
<td>0.25</td>
<td>0.75</td>
</tr>
<tr>
<td>High partition</td>
<td>.32</td>
<td>.70</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>NASDAQ “nonexperts”</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low partition</td>
<td>0.80</td>
<td>0.75</td>
</tr>
<tr>
<td>High partition</td>
<td>0.12</td>
<td>.88</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>22</td>
</tr>
</tbody>
</table>

$nasdaq$"experts"

$nasdaq$"nonexperts"
Table 4. Median Probabilities from Study 5. “Low Partition” refers to the partition with intervals 400 or less, 401-600, 601-800, 801-1000, and above 1000. “High Partition” refers to the partition with intervals 1000 or less, 1001-1200, 1201-1400, 1401-1600, and above 1600. Thus, the numbers shown in bold italics result from combining separate assessments for four intervals. Columns labeled “n” indicate the number of usable responses for the corresponding cell.

<table>
<thead>
<tr>
<th></th>
<th>DAS</th>
<th></th>
<th>SQA</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>≤1000</td>
<td>&gt;1000</td>
<td>n</td>
<td>≤1000</td>
</tr>
<tr>
<td>Low Partition</td>
<td>0.90</td>
<td>0.10</td>
<td>29</td>
<td>0.80</td>
</tr>
<tr>
<td>High Partition</td>
<td>0.65</td>
<td>0.35</td>
<td>28</td>
<td>0.52</td>
</tr>
</tbody>
</table>

Table 5. Median Probabilities for 25 Most Experienced Decision Analysts in Study 5. Although there were a total of 25 respondents included in this subsample, only 12 out of 13 provided usable responses for the DAS low partition and for the SQA high partition.

<table>
<thead>
<tr>
<th></th>
<th>DAS</th>
<th></th>
<th>SQA</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>≤1000</td>
<td>&gt;1000</td>
<td>n</td>
<td>≤1000</td>
</tr>
<tr>
<td>Low Partition</td>
<td>0.88</td>
<td>0.12</td>
<td>12</td>
<td>0.80</td>
</tr>
<tr>
<td>High Partition</td>
<td>0.73</td>
<td>0.27</td>
<td>12</td>
<td>0.58</td>
</tr>
</tbody>
</table>
Figure 1. Possible Reasons Why a Car Might Fail to Start. Fischhoff, Slovic, and Lichtenstein (1978) showed one group of participants the upper tree and another group the lower tree and asked participants to estimate the number of times out of 1000 that a car would fail to start for a reason contained in each of the main categories (e.g., “Starting system defective,” “Ignition system defective,” etc.). Although the categories “Battery charge insufficient,” “Fuel system defective,” and “Other engine problems” were relegated to “All other problems” in the lower tree, the judged probability of “All other problems” in the lower tree did not increase by a corresponding amount.
Figure 2. Results from Study 2. The median judgment of $P(\text{JSX} < \$1000)$ in the three-fold low condition is about twice what it is in six-fold low. Conversely, $P(\text{JSX} \geq \$1000)$ in the three-fold high condition is about twice what it is in the six-fold high condition. In all cases, the median probability is similar to the ignorance prior.

![Graph showing results from Study 2.](image)

Legend: □ Median  
| Upper and lower quartiles  
◆ Ignorance prior

Figure 3. Stimuli for Study 3. Participants were asked to assess the probability that the starting salary for a randomly chosen member of the next graduating class would fall into each interval. (The dashed lines were not included in the questionnaire and are shown here solely to clarify the experimental design.) Each participant responded either to both low partitions or to both high partitions. The partitions for Duke MBA and Harvard Law differ slightly reflecting the results of a pretest that provided rough interval estimates of the two quantities.

<table>
<thead>
<tr>
<th>Low Partitions</th>
<th>Duke MBA</th>
<th>Harvard Law</th>
</tr>
</thead>
<tbody>
<tr>
<td>$55,000 or less</td>
<td>____ %</td>
<td>$60,000 or less</td>
</tr>
<tr>
<td>$55,001-$65,000</td>
<td>____ %</td>
<td>$60,001-$70,000</td>
</tr>
<tr>
<td>$65,001-$75,000</td>
<td>____ %</td>
<td>$70,001-$80,000</td>
</tr>
<tr>
<td>$75,001-$85,000</td>
<td>____ %</td>
<td>$80,001-$90,000</td>
</tr>
<tr>
<td>More than $85,000</td>
<td>____ %</td>
<td>More than $90,000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>High Partitions</th>
<th>Duke MBA</th>
<th>Harvard Law</th>
</tr>
</thead>
<tbody>
<tr>
<td>$85,000 or less</td>
<td>____ %</td>
<td>$90,000 or less</td>
</tr>
<tr>
<td>$85,001-$95,000</td>
<td>____ %</td>
<td>$90,001-$105,000</td>
</tr>
<tr>
<td>$95,001-$105,000</td>
<td>____ %</td>
<td>$105,001-$115,000</td>
</tr>
<tr>
<td>$105,001-$115,000</td>
<td>____ %</td>
<td>$115,001-$130,000</td>
</tr>
<tr>
<td>More than $115,000</td>
<td>____ %</td>
<td>More than $130,000</td>
</tr>
</tbody>
</table>
Figure 4. Stimuli for Study 4. This experiment was run in July, 2000. The NASDAQ index was near 4000 at the time.

1) What is the last digit of your local telephone number? ________

If this number is *even*, please write “JSX” in the space provided above the tree on the *left* and “NASDAQ” in the space provided above the tree on the *right*.

If this number is *odd*, please write “NASDAQ” in the space provided above the tree on the *left* and “JSX” in the space provided above the tree on the *right*.

<table>
<thead>
<tr>
<th>Index: _______________</th>
<th>Index: _______________</th>
</tr>
</thead>
<tbody>
<tr>
<td>Below 1000</td>
<td>4000 or below</td>
</tr>
<tr>
<td>1000-2000</td>
<td>4001-8000</td>
</tr>
<tr>
<td>2001-4000</td>
<td>8001-16000</td>
</tr>
<tr>
<td>Above 4000</td>
<td>Above 16000</td>
</tr>
</tbody>
</table>

2) For each tree above please estimate the probabilities that the designated index will close in each specified range on the last day of trading this year. Be sure that the four probabilities for a given index sum to 100 percent.

3) Please rate your familiarity with each of the two indices on a 0-10 scale (0 = I know nothing; 10 = I know it extremely well) by placing a number beside each index name that you wrote above.
Figure 5. Stimuli for Study 5. In the questionnaire, DAS refers to the Decision Analysis Society and SQA to the Society of Quantitative Analysts, a professional society incorporated in 1989 that is concerned with the application of quantitative techniques in finance, investment, and risk management. Although not indicated here, we also asked about the individual’s level of education, whether he or she had taught decision analysis, the number of applied projects over the past two years in which he or she had elicited probabilities, and elicitation procedures used.

Is the last digit of your primary home telephone number even or odd? __

IF IT IS EVEN, type "DAS" below in the space following "SOCIETY 1" and type "SQA" in the space following "SOCIETY 2."

IF IT IS ODD, type "SQA" in the space following "SOCIETY 1" and type "DAS" in the space following "SOCIETY 2."

SOCIETY 1: ___________________

Please assess your subjective probability that the total membership for this society five years from today will fall in the indicated interval. Please be sure that your probabilities add up to 1.00:

P(membership 400 or less) = __________
P(membership between 401 and 600) = __________
P(membership between 601 and 800) = __________
P(membership between 801 and 1000) = __________
P(membership more than 1000) = __________

SOCIETY 2: ___________________

Please assess your subjective probability that the total membership for this society five years from today will fall in the indicated interval. Please be sure that your probabilities add up to 1.00:

P(membership 1000 or less) = __________
P(membership between 1001 and 1200) = __________
P(membership between 1201 and 1400) = __________
P(membership between 1401 and 1600) = __________
P(membership more than 1600) = __________