DRAFT: 7/6/04 Comments welcome

Assigning subjective probabilities to event trees: 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

Address Correspondence to:
Robert T. Clemen
Fuqua School of Business
Duke University, Box 90120
Durham, NC 27708-0120
919-660-8005
clemen@duke.edu

Assigning subjective probabilities to event trees:
Partition dependence and bias toward the ignorance prior

Abstract

Decision and risk analysts have considerable discretion in designing procedures for eliciting subjective probabilities. One popular approach is to specify a particular set of exclusive and exhaustive events for which the assessor provides subjective probabilities. We show that assessed probabilities are biased toward a uniform distribution over all events into which the relevant sample space happens to be partitioned. We surmise that an assessor begins with an "ignorance prior" probability distribution that assigns equal probabilities to the specified events, then insufficiently adjusts those probabilities to reflect his or her beliefs concerning how the likelihood 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 (but may or may not disappear) 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 (enhanced availability of specified events, ambiguity in interpreting event categories, demand effects) cannot fully account for the effect. 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 probability-assessment protocols described more recently by Keeney and von Winterfeldt (1991) and Morgan and Henrion (1990).

Human limitations of attention, memory, and information processing capacity can lead to subjective probabilities that are often 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), relatively little attention is given to the choice of specific events to be assessed.

In developing an elicitation structure, the analyst chooses the frame within which the expert constructs 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, the analyst might ask an expert to assess probabilities that next quarter's sales will rise or fall relative to their current level; another expert might be asked to assess the probability that completion time for a project will exceed the budgeted time. Intervals may also be dictated by thresholds relevant to the decision. For instance, a farmer might assess the probability that next year's crop price will exceed 38 cents per bushel because 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 affect assessed probabilities. Unfortunately, our experimental results demonstrate that this assumption is unfounded: assessed probabilities can depend crucially on 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 in the seminal paper 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 categories are singled out and made explicit rather than lumped into a catch-all category, people can more readily retrieve instances of these categories from memory and therefore judge these categories to be more likely.

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 cast doubt on the generality of the traditional availability-based account and instead attribute pruning bias—and more generally partition dependence—to a tendency to first allocate probabilities equally over all exclusive and exhaustive events that have been identified and then insufficiently adjust this uniform distribution according to beliefs about how the likelihood of the events differ. Understanding the nature and causes of partition dependence can help analysts identify conditions under which this bias may arise, anticipate specific patterns of bias, predict conditions that may exacerbate or mitigate the effect, and develop useful debiasing techniques.

In the following section of this paper we review literature on the pruning bias and partition dependence. In §3 we describe a series of studies that document partition dependence in a variety of contexts and provide support for our interpretation of this phenomenon. We close with a discussion of the interpretation and robustness of partition dependence, other manifestation of this phenomenon, and the prescriptive implications of the present results.

2. Literature Review and Theory

FSL presented professional automobile mechanics and laypeople with trees that identified several specific reasons why a car might fail to start as well as a residual catch-all category of reasons labeled “all other problems.” Participants were asked to estimate the number of times out of 100 that a car would fail to start for each of the reasons specified. When the experimenters removed (*pruned*) specific 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 a group of participants, did not increase by a corresponding amount. Instead, the probability for the category that was 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). In explaining this phenomenon, FSL invoked the availability heuristic (Tversky & Kahneman, 1973), whereby frequency or probability of events is judged by the ease with which instances can be brought to mind. Because it is easier to recruit instances of a particular reason why a car might not start when the reason is explicitly stated, probabilities are higher for causes that are explicitly identified than for unstated causes that are implicitly included in a residual category.

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

Availability. Van der Pligt, Eiser, and Speark (1987) replicated the pruning bias in a within-subject study of British voters who were asked on multiple occasions to estimate the percentage of power that is provided by various sources, with both pruned and unpruned trees presented to the same participants on different occasions. Like FSL, these researchers interpreted their result in terms of the availability heuristic.

Dubé-Rioux and Russo (1988) provided more direct evidence for the availability hypothesis by asking managers in the hospitality industry to assign probabilities to various possible reasons for a decrease in restaurant profits. At various points in the study, participants were presented with (1) a full tree listing 12 specific causes and a residual category, (2) a pruned tree listing 6 specific causes and a residual category, and (3) an "extended-pruned" trees in which 6 specific causes were listed, but participants were prompted to add others to the list before assigning probabilities to their elaborated tree. This latter manipulation reduced the magnitude of the pruning bias; moreover, the pruning bias was smaller among participants who were able to generate a greater number of additional causes. These results suggest that the accessibility of a particular category mediates the extent of the pruning bias.

Perhaps the most direct evidence for the availability interpretation comes from Ofir (2000), who argued that people should rely more heavily on the availability heuristic when they know more about the events for which probabilities are being assessed. Ofir noted that the original characterization of the availability heuristic 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). In one study, auto mechanics and laypersons with driver licenses were asked to judge the probabilities of various causes of a car not starting. Participants were presented with pruned versions of the FSL tree and asked to generate either one or four specific causes that would fall under the residual category, then judge the probabilities of all branches of the tree (treating the elaborated residual as a single category). The lay drivers found it relatively easy to generate one cause for the residual category but relatively difficult to generate four specific causes; they actually assigned a *lower* probability when asked to generate four causes than when they were asked to generate a single cause. The mechanics also found it more difficult to generate a larger number of causes, but unlike the lay drivers they assigned a higher probability when asked to generate four causes. These results suggest that people with less domain knowledge rely on the ease with which they can recall instances (i.e., the availability heuristic), whereas people with more domain knowledge are influenced by the absolute number of instances that come to mind.

Ambiguity. Hirt and Castellan (1988) argued that the 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 suggests that 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-Rioux & Russo, 1988; see also FSL, p. 340). Reasonable probability assessors may take for granted that they have been presented with a credible fault tree, and they may therefore infer that the residual category should have a relatively low probability of occurrence. 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 the tree is a credible representation of possible causes of failure so that each branch to be assessed should be assigned a nontrivial probability. We note that the results from FSL's Experiment 1 call the credibility account into question, because the mean probability assigned to the least important of seven branches was only 0.033.

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: FSL's reasons for a car not starting and causes of death for a randomly selected individual. 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 observed 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, 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. In one condition participants were told that 15-year-old students had generated the categories. Pruning bias was not significantly larger in this condition, suggesting that the credibility of the source of an event tree does not influence participants' tendency to spread probabilities evenly among the branches of that tree.

Anchoring and insufficient adjustment. A fourth interpretation of the pruning bias is the notion that 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. Anchoring 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 "principle of insufficient reason," advanced by early probability theorists such as Leibniz and Laplace, who argued that if there are no obvious reasons why one event should be more likely than others, then these events should be treated as equally likely. 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 existing empirical evidence for this anchoring hypothesis 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 1) asked participants “what should be” the probability of a residual category for trees with different numbers (n) of labeled branches; the responses provided a “remarkable fit” to the formula $p_n=1/(n+1)$.

In the section that follows we offer more direct evidence that pruning bias is driven primarily by a tendency to allocate probability evenly across all events into which the event space happens to be partitioned (including the residual event). 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 exhibit partition dependence in situations where the availability, ambiguity, and credibility mechanisms are not expected to play a significant role.

3. Experimental Evidence

Study 1: Unpacking versus separate evaluation

To disentangle the availability account from anchoring on the ignorance prior, note that 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, we predict that the distribution of probabilities will be affected primarily by the number of branches that are explicitly evaluated. Indeed, some studies suggest that events are generally assigned higher probabilities when they are split into multiple branches that are evaluated separately. In FSL's Experiment 5, the authors either split a single branch of a base fault tree into two branches (e.g. by splitting off "distribution system defective" into a separate category from "ignition system defective"), or fused two branches in the base tree into a single branch (e.g., by combining "starter system defective" and "ignition system defective"). In both the split and fused trees the same events were explicitly identified, but the grouping of events to be evaluated differed by experimental condition. The sum of the probabilities of events in the split tree tended to be higher than the probabilities of corresponding events of the base tree, while the sum of probabilities of events in the base tree tended to be higher than the probabilities of corresponding events in the fused tree. A similar pattern was observed by Van Schie and Van der Pligt (1990) when they asked students to estimate the proportion of acid rain that could be attributed to various factors.

In support theory, Rottenstreich and Tversky (1997) provided evidence that (1) unpacking a category (e.g., homicide) into a disjunction of subcategories (e.g., homicide by an acquaintance or homicide by a stranger) increases judged probability, and (2) separate assessment of subcategories increases aggregate judged probabilities still further. A recent review of the support theory literature by Sloman et al. (2004) suggests that the second pattern is more robust and more pronounced than the first pattern. In fact, these authors provide several new examples that violate the first pattern. The more robust separate-evaluation effect is consistent with a bias toward 1/*n* for each of *n* exclusive and exhaustive events that are evaluated in parallel.

Our first study was designed to demonstrate the effects of separate evaluation in a context where these effects cannot readily be attributed to the availability or ambiguity explanations. 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 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. In a survey of students admitted to Duke's daytime MBA program (N = 285), 99% of respondents indicated that they had used *Business Week* and/or *US News and World Report's* published rankings of business schools in making their decisions about which business school to attend. Participants 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 same schools were listed alphabetically on separate branches:

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

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 four of the strongest competitors to Wharton:

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

Finally, participants in the *pruned-tree* condition (n = 31) were presented a tree with the following branches:

- A school other than Wharton
- Wharton

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 Rottenstriech & 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 availability-based accounts are not a necessary source of the pruning bias. In both the implicit (*pruned*) and explicit (*collapsed*) conjunctions 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: Partition dependence under ignorance

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 (χ^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 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. The second domain was the starting salary of Harvard Law graduates, about which we expected our participants to know relatively little.

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 posed the problem of judging the probability 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^{th}, 50^{th}, and 90^{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 corresponding 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. Table 2 presents results from Study 3. 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.

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_{high}(Harvard<90K) denote the single judged probability that a randomly chosen graduate of Harvard Law will earn less than $90,000 during his or her first year after graduation (the top entry in the lower right-hand cell in Table 2). Let P_{low}(Harvard<90K) denote the corresponding sum of judged probabilities for a participant in the *low* partition (the sum of the top four entries in the upper right-hand cell in Table 2). Finally, define P_{high}(Duke<85K) and P_{low}(Duke<85K) similarly. Median probabilities presented in Table 2 reveal partition dependence for both Harvard and Duke. In particular, judged probabilities are lower when they are derived from a single judgment than when they are derived from multiple judgments that are summed: P_{high}(Harvard<90K) < P_{low}(Harvard<90K) and P_{high}(Duke<85K) < P_{low}(Duke<85K). To perform an overall test for significance of the effect, we calculated P_{low}(Harvard<90K) + P_{low}(Duke<85K) for each participant in the *low* condition, and P_{high}(Harvard<90K) + P_{high}(Duke<85K) for each participant in the *high* condition. The ignorance-prior account predicts that the median sum for the *low* participants will be greater than the median sum for the *high* participants; this prediction is confirmed by a one-tailed Wilcoxon rank-sum test ($p < .0001$).

As predicted, partition dependence is dramatic for Harvard (difference of medians = 0.45) and less pronounced for Duke (difference of medians = 0.35). To test the statistical significance of this pattern, we calculated P_{low}(Harvard<90K) - P_{low}(Duke<85K) among participants in the low partition conditions (for whom these values were the sums of four separate judgments) and P_{high}(Harvard<90K) - P_{high}(Duke<85K) for 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 approaches significance by a one-tailed Wilcoxon test (p =.12) and by t-test ($t(105) = 1.63$, $p = .05$).

Study 4: Ruling out credibility and demand effects

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 assume that they have been provided with a credible event tree in which each 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. To rule out this account, 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 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 in the questionnaire. Two different 4-interval partitions were specified: 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 variable 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 randomly to partition conditions. If there were any information to be gleaned from the particular partitions, it might have been that the 4000 threshold was particularly relevant. (At the time of the study the NASDAQ index was closing near 4000.) There was no reason, however, to believe that the condition assignment should affect judged probabilities.

Results and discussion. Table 3 summarizes the results from 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* expectations. 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 from participants whose probabilities 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 each interval for JSX was 25% in both *low* and *high* partition conditions. More strikingly, the 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 conducted a median split on knowledge ratings for the NASDAQ. Table 3 shows the median 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 eliminated in the NASDAQ assessments for the experts, (difference in medians = -0.07, p = .44), and the interaction is highly significant (p = .001).[1] There is no difference, however, between NASDAQ experts and non-experts on the JSX assessments. These results provide further support for the claim that increased knowledge leads to a decrease in partition dependence. 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 in FSL, and Ofir (2000)).

Study 5: Experts in probability assessment

The results of Studies 1-4 demonstrate the robustness of partition dependence among a fairly sophisticated population: graduate students of business, some 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 these studies did not have extensive training in and experience using probability-assessment methods, and that

---

[1] This statistic was computed by comparing pooled expert-packed and nonexpert-unpacked groups with pooled nonexpert-packed and expert-unpacked groups (i.e., comparing the diagonal with off-diagonal cells) using a Kruskal-Wallis test.

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.

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," although 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) and, if PhD, what year; whether he or she had taught a course in decision analysis and the most recent year in which this had occurred; 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 ($p = .45$, one-tailed), though it was significant by a parametric test ($t(44) = -2.28$, $p = .01$).

To explore the robustness of our results we examined a subsample of 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 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 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 stems from details of the task: 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. This combination may have given rise to similar knowledge concerning the two assessment tasks for our participants.

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 evidence for events that happen to be made explicit or ambiguity of event categories. In Study 1 we showed that unpacking 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. This tendency was also reflected in betting behavior. Study 3 demonstrated that partition dependence among Duke MBA students was more pronounced for a less familiar domain (Harvard Law graduate salaries) than a more 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 on the part of the 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 (ignorance-prior) probabilities for each event in the specified partition 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 events are driven largely by anchoring on the ignorance prior, though we acknowledge that availability effects may also contribute.

Partition dependence appears to be robust across a variety of substantive contexts in which availability, credibility, and ambiguity provide unsatisfactory explanations. Consistent with our anchoring-and-adjustment account, partition dependence is reduced among participants with greater substantive expertise and may disappear in situations where participants are especially knowledgeable (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 MBAs 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 demonstrated a subjective feeling of relatively high expertise.[2] 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

[2] Partition dependence was also observed among experts concerning MBA program rankings in Study 1 and among experienced auto mechanics in the studies of FSL and Ofir (2000), although we note that these studies employed partitions that entailed discrete events. One might speculate that partition dependence could be more robust to expertise in such situations where both availability and a bias toward the ignorance prior may affect judgment.

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 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. Finally, Fox and Levav (2004) showed that when people have difficulties solving conditional probability puzzles such as the "Monty Hall" problem, those difficulties may reflect naïve extensional reasoning in which they subjectively partition the sample space on the basis of initial conditions, edit the partition using conditioning information, and calculate probability on the basis of the ratio of remaining events in the partition. They show further that subtle rewording of these problems can enhance the use of more refined 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 two-fold partition {other child boy, other child girl} and an incorrect response of 1/2. Very few participants provided the correct response of 1/3. However, the informationally 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 {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. For example, Birnbaum and colleagues have studied event-splitting effects on risky choice, showing that manipulation of the event partition can dramatically affect the tendency of participants to choose a stochastically dominated lottery (for a review see Birnbaum & Martin, 2003).

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 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 lotteries, the units that are allocated (dollars versus number of shares), and the procedure used to indicate preferences (explicit allocation versus choice of most preferred portfolio). 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).

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 such as those used in this study (where intervals are either provided by the analyst or suggested by the expert). Another predominant technique is to elicit fractiles for the uncertain variable (often 10^{th}, 50^{th}, and 90^{th} 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 10^{th} percentile or above the 90^{th} 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 partition dependence plays a role in fractile elicitation.

Partition dependence has clear implications for decision and risk analysis. Because expert probability assessments can depend strongly on the specific partition used, at a minimum the analyst should strive to direct the expert's attention across the event space in as evenhanded a way as possible, beginning with a careful specification of the partition. 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 harder to compare some partitions in terms of evenhandedness. For instance, consider a 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°C $< T \leq 5^\circ$C, 5°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 continuous variables.

Although our results lead directly to the advice to specify partitions with care, a deeper understanding of both the psychological processes underlying partition dependence and how debiasing approaches may affect these processes can help us to identify additional procedures for counteracting partition dependence (Fischhoff, 1982). We speculate that partition dependence arises from an automatic, non-conscious process whereby the judge begins with the default ignorance prior (or equal credence in all identified events) and then adjusts his or her probabilities (insufficiently) to account for relevant knowledge. We surmise that the specification of particular events in a partition may make related events seem more plausible by priming semantically related information. Arkes (1991) reviews evidence showing that such *association-based* judgment errors are quite resistant to warnings or stronger incentives, though some biases (e.g. hindsight bias, overconfidence) can be somewhat mitigated by asking the expert to perform specific, compensating behaviors. Kahneman and Frederick (2002) and Larrick (in press) suggest that errors arising from automatic processes can sometimes be debiased using deliberate, conscious reasoning—but only to the extent that the expert recognizes an error or inconsistency. We conclude from this literature that debiasing may be possible through some sort of compensating procedure that calls attention to inconsistency or error in an expert's judgment.

Probability-elicitation procedures used in decision and risk analysis (Clemen, 1996; 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. An example is the use of the probability wheel as a "thought experiment" to aid probability elicitation; the analyst can use the wheel to encourage the expert to think carefully about the uncertain variable in question. Counteracting partition dependence requires thorough and even-handed exploration of the event space. Best practice in probability elicitation may help in this regard; we paraphrase such advice (described 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^{th} and 90^{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 sample 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. Exactly how an analyst should develop multiple partitions will naturally vary case-by-case. Thus, it may make sense for the analyst and expert to work together to specify different partitions and compare results.

The motivation for using multiple partitions is the same as the motivation for using multiple assessment methods: the analyst can help the expert encode his or her beliefs carefully and deliberately, avoiding various judgmental pitfalls including partition dependence and arriving at a set of consistent probability judgments. We note, though, 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 average 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. If our suggested approach helps experts encode their knowledge with greater care, improvements in calibration, resolution, or both might be expected.

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

Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin, 110*, 486-498.

Benartzi, S., & Thaler, R. (2001). Naive diversification strategies in retirement saving plans. *American Economic Review, 91*, 79-98.

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.

Brier, G. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78, 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.

Clark, H. H. (1985). Language use and language users. In: G. Lindzey & E Aronson (Eds.) *Handbook of social psychology, Vol 2* (pp. 179-232). New York: Random House.

Clemen, R. T. (1996). *Making hard decisions: An introduction to decision analysis, 2nd Ed*. Belmont, CA: Duxbury.

Dubé-Rioux, L., & Russo, J. E. (1988). An availability bias in professional judgment. *Journal of Behavioral Decision Making, 1*, 223-237.

Epley, N. & Gilovich, T. (2001). Putting adjustment back into the anchoring-and-adjustment heuristic: Self-generated versus experimenter provided anchors. *Psychological Science, 12,* 391-396.

Fiedler, K., & Armbruster, T. (1994). Two halfs may be more than one whole: Category-split effects on frequency illusions. *Journal of Personality and Social Psychology, 66,* 633-645.

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

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1978). Fault trees: Sensitivity of estimated failure probabilities to problem representation. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 330-344.

Fox, C. R., Bardolet, D., & Lieb, D. (in press). Partition dependence in decision analysis, resource allocation, and consumer choice. Chapter to appear in R. Zwick & A. Rappoport (Eds.) *Experimental business research, Vol. 3.* Dordrecht, NL: Kluwer.

Fox, C. R., & Levav, J. (in press). Partition-edit-count: Naïve extensional reasoning in judgment of conditional probabilities. *Journal of Experimental Psychology: General,* forthcoming.

Fox, C. R., & Rottenstreich, Y. (2003). Partition priming in judgment under uncertainty. *Psychological Science, 13,* 195-200.

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.) (2002). *Heuristics and biases: The psychology of intuitive judgment.* Cambridge, UK: Cambridge University Press.

Grice, H. P. (1975). Logic and conversation. In: P. Cole & J. L. Morgan (Eds.) Syntax and Semantics, Vol. 3, Speech Acts. New York: Academic Press, 41-58.

Harries, C., & Harvey, N. (2000). Are absolute frequencies, relative frequencies, or both effective in reducing cognitive biases? *Journal of Behavioral Decision Making, 13*, 431-444.

Hirt, E. R., & Castellan, N. J., Jr. (1988). Probability and category identification in the fault tree paradigm. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 112-131.

Johnson, R.D., Rennie, R.D., & Wells, G.L. (1991). Outcome trees and baseball: A study of expertise and list-length effects. *Organization Behavior and Human Decision Processes, 50,* 324-340.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In: T. Gilovich, D. Griffin, & D. Kahneman (Eds.) *Heuristics and biases: The psychology of intuitive judgment* (pp. 49-81). Cambridge, UK: Cambridge University Press.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.) (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, UK: Cambridge University Press.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist, 39*, 341-50.

Keeney, R., & von Winterfeldt, D. (1991). Eliciting probabilities from experts in complex technical problems. *IEEE Transactions on Engineering Management, 38*, 191-201.

Klayman, J., Soll, J.B., Gonzales-Vallejo, C., & Barlas, S. (1999). Overconfidence: It depends on how, what, and whom you ask. *Organizational Behavior and Human Decision Processes, 79,* 216-247.

Langer, T., & Fox, C. R. (2004). Naïve diversification in allocation among risky prospects: Partition-dependence, unit-dependence, and procedure-dependence. Unpublished manuscript, University of Mannheim.

Larrick, R. P. (in press). Debiasing. In: D. Koehler & N. Harvey (Eds.) *The Blackwell handbook of judgment and decision making*. Oxford: Blackwell, forthcoming.

Lichtenstein, S., Fischhoff, B., & Phillips, L. (1982). Calibration of probabilities: The state of the art to 1980. In: D. Kahneman, P. Slovic, & A. Tversky (Eds.) *Judgment under uncertainty: Heuristics and biases* (pp. 306-334). Cambridge, UK: Cambridge University Press.

Morgan, M., & Henrion, M. (1990). *Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy management*. Cambridge, UK: Cambridge University Press.

Ofir, C. (2000). Ease of recall vs recalled evidence in judgment: Experts vs laymen. *Organizational Behavior and Human Decision Processes, 81*, 28-42.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17*, 776-783.

Rottenstreich, Y., & Tversky, A. (1997). Unpacking, repacking, and anchoring: Advances in support theory. *Psychological Review, 2*, 406-415.

Russo, J. E., & Kolzow, K. J. (1994). Where is the fault in fault trees? *Journal of Experimental Psychology: Human Perception and Performance, 20*, 17-32.

Schwarz, N., Bless, B., Strack, G., Klumpp, G., Schatka, H.R., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology, 61*, 195-202.

See, K.E., Fox, C.R., & Rottenstreich, Y. (2003). Attribute-cued partitioning and learning in judgment under uncertainty. Unpublished manuscript, Wharton School of Business, University of Pennsylvania.

Sloman, S., Rottenstreich, Y., Wisniewski, E., Hadjichristidis, C., & Fox, C.R. (2004). Typical versus atypical unpacking and superadditive probability judgment. *Journal of Experimental Psychology: Learning, Memory & Cognition, 30,* 573-582.

Spetzler, C. S., & Staël Von Holstein, C.-A. (1975). Probability encoding in decision analysis. *Management Science*, *22*, 340-352.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology. 4*, 207-232.

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

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business*, *59*, S251-S278.

Van der Pligt, J., Eiser, J.R., & Speark, R. (1987). Comparative judgments and preferences: The influence of the number of response alternatives. *British Journal of Social Psychology, 26,* 269-280.

Van Schie, E.C.M., & Van der Pligt, J. (1990). Influence diagrams and fault trees: The role of salience and anchoring. In K. Borsherding, O.I. Larichev, & D.M. Messick (Eds.), *Contemporary Issues in Decision Making* (pp. 177-188). Amsterdam: North-Holland.

Van Schie, E. C. M., & Van der Pligt, J. (1994). Getting an anchor on availability in causal judgment. *Organizational Behavior and Human Decision Processes*, *57*, 140-154.

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

Weber, M., Eisenfuhr, F., & von Winterfeldt, D. (1988). The effects of splitting attributes on weights in multiattribute utility measurement. *Management Science, 34*, 431-445.

Yates, J. F. (1990). *Judgment and decision making*. Englewood Cliffs, NJ: Prentice-Hall.

Table 1. Study 1 Results

Condition	Median P(School other than Wharton is ranked #1)	Significance levels using Wilcoxon rank-sum test (one-tailed)	
		Comparing	p-value
Pruned	0.40	*Pruned* vs. *Collapsed*	.350
Collapsed	0.40	*Pruned* vs. *Full Tree*	.005
Full tree	0.70	*Collapsed* vs. *Full Tree*	.053

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.

	Duke MBA			**Harvard Law**		
Knowledge Rating	7/10			2/10		
	≤ $85,000	> $85,000	*n*	≤ $90,000	> $90,000	*n*
Low Partition	***0.75***	0.25	55	***0.75***	0.25	57
High Partition	0.40	***0.60***	57	0.30	***0.70***	58

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.

		NASDAQ			**JSX**		
Knowledge rating		7/10			0/10		
		≤4000	>4000	*n*	≤4000	>4000	*n*
Full data set	*Low* partition	***0.50***	0.50	49	***0.75***	0.25	52
	High partition	0.25	***0.75***	51	0.25	***0.75***	45
NASDAQ "experts"	*Low* partition	***0.25***	0.75	24	***0.75***	0.25	24
	High partition	0.32	***0.70***	24	0.25	***0.75***	23
NASDAQ "nonexperts"	*Low* partition	***0.80***	0.20	21	***0.75***	0.25	22
	High partition	0.12	***0.88***	22	0.23	***0.77***	18

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.

	DAS			**SQA**		
	≤1000	>1000	*n*	≤1000	>1000	*n*
Low Partition	***0.90***	0.10	29	***0.80***	0.20	27
High Partition	0.65	***0.35***	28	0.52	***0.48***	28

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.

	DAS			**SQA**		
	≤1000	>1000	*n*	≤1000	>1000	*n*
Low Partition	***0.88***	0.12	12	***0.80***	0.20	12
High Partition	0.73	***0.27***	12	0.58	***0.42***	12

Figure 1. Possible Reasons Why a Car Might Fail to Start. Fischhoff, Slovic, and Lichtenstein (1978) showed one group of participants the tree on the left and another group the tree on the right and asked participants to estimate the number of times out of 100 that a car would fail to start for each of the reasons specified. Although the cause "fuel system defective" was relegated to "all other problems" in the tree on the right, the judged probability of "all other problems" did not increase by a corresponding amount.

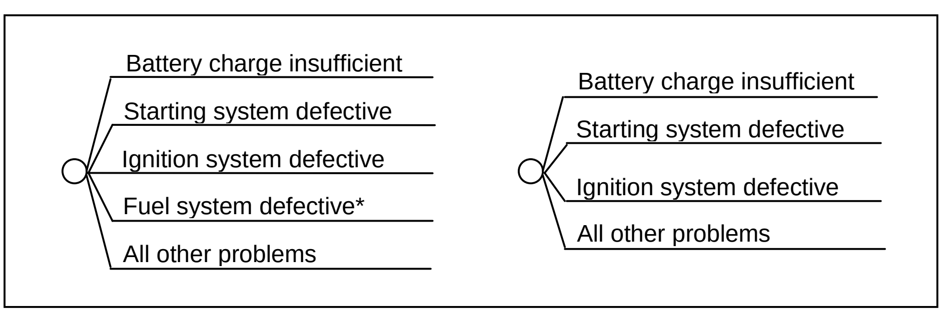

Figure 2. Results from Study 2. The median judgment of P(JSX < $1000) in the *three-fold low* condition is about twice what it is in *six-fold low*. Conversely, P(JSX ≥ $1000) in the *three-fold high* condition is about twice what it is in the *six-fold high* condition. In all cases, the median probability is similar to the ignorance prior.

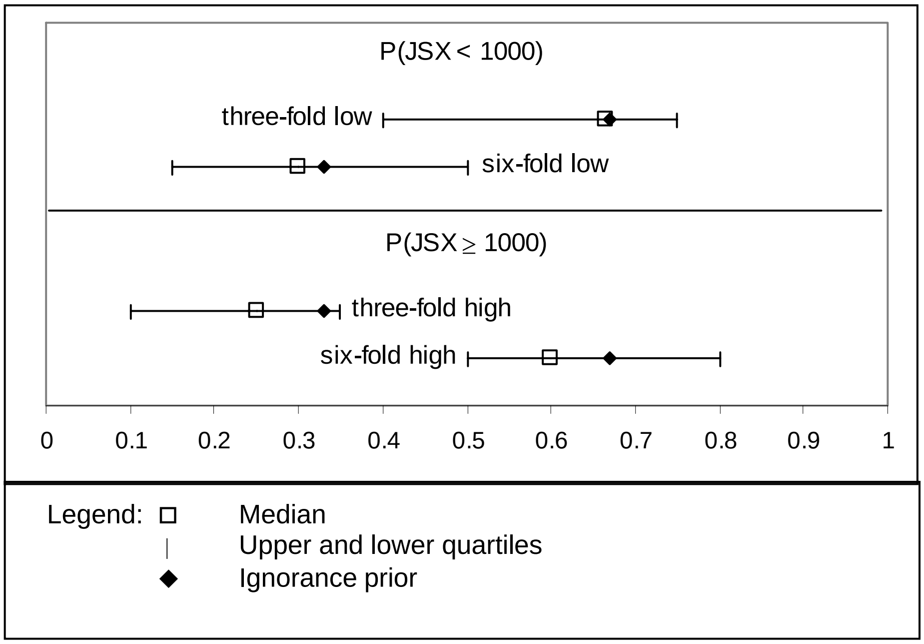

Figure 3. Stimuli for Study 3. Participants were asked to assess the probability that the starting salary for a randomly chosen member of the next graduating class would fall into each interval. (The dashed lines were not included in the questionnaire and are shown here solely to clarify the experimental 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.

	Duke MBA		**Harvard Law**	
Low Partitions	$55,000 or less	______%	$60,000 or less	______%
	$55,001-$65,000	______%	$60,001-$70,000	______%
	$65,001-$75,000	______%	$70,001-$80,000	______%
	$75,001-$85,000	______%	$80,001-$90,000	______%
	- - - - - - - - - - - - - - - -		- - - - - - - - - - - - - - - -	
	More than $85,000	______%	More than $90,000	______%
High Partitions	$85,000 or less	______%	$90,000 or less	______%
	- - - - - - - - - - - - - - - -		- - - - - - - - - - - - - - - -	
	$85,001-$95,000	______%	$90,001-$105,000	______%
	$95,001-$105,000	______%	$105,001-$115,000	______%
	$105,001-$115,000	______%	$115,001-$130,000	______%
	More than $115,000	______%	More than $130,000	______%

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.*

Index:____________________		Index: __________________	
Below 1000	______	4000 or below	______
1000-2000	______	4001-8000	______
2001-4000	______	8001-16000	______
Above 4000	______	Above 16000	______

2) For each tree above please estimate the probabilities that the designated index will close in each specified range <u>on the last day of trading this year</u>. 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) = ___________