MANAGEMENT SCIENCE
Vol. 54, No. 4, April 2008, pp. 835–851
ISSN 0025-1909 | EISSN 1526-5501 | 08 | 5404 | 0835
DOI 10.1287/mnsc.1070.0790


# Interior Additivity and Subjective Probability Assessment of Continuous Variables

Robert T. Clemen
Fuqua School of Business, Duke University, Durham, North Carolina 27708, clemen@duke.edu

Canan Ulu
McCombs School of Business, University of Texas at Austin, Austin, Texas 78712,
canan.ulu@mccombs.utexas.edu

One of the goals of psychological research on subjective probability judgment is to develop prescriptive procedures that can improve such judgments. In this paper, our aim is to reduce partition dependence, a judgmental bias that arises from the particular way in which a state space is partitioned for the purposes of making probability judgments. We explore a property of subjective probabilities called interior additivity (IA). Our story begins with a psychological model of subjective probability judgment that exhibits IA. The model is a linear combination of underlying support for the event in question and a term that reflects a prior belief that all elements in the state space partition are equally likely. The model is consistent with known properties of subjective probabilities, such as binary complementarity, subadditivity, and partition dependence, and has several additional properties related to IA. We present experimental evidence to support our model. The model further suggests a simple prescriptive method based on IA that decision and risk analysts can use to reduce partition dependence, and we present preliminary empirical evidence demonstrating the effectiveness of the method.

*Key words*: decision analysis; subjective probability assessment; support theory; partition dependence; ignorance prior; debiasing
*History*: Accepted by George Wu, decision analysis; received March 27, 2006. This paper was with the authors $7\frac{1}{2}$ months for 2 revisions.

## 1. Introduction

To quantify uncertainty, decision and risk analysts make extensive use of subjective probability judgments. Such judgments can be crucial in many decision situations, such as siting a nuclear waste repository or choosing an optimal portfolio of research and development projects. Extensive study of subjective probability judgments, however, has shown that these judgments—even those of experts—often display a variety of biases (see, e.g., Kahneman et al. 1982, Gilovich et al. 2002). Analysts typically work hard to reduce biases when eliciting experts' subjective beliefs. Spetzler and Staël Von Holstein (1975) laid out practical elicitation methods that were designed specifically to counteract a variety of biases. Spetzler and Staël Von Holstein's (1975) methods are still in use, as evidenced by their incorporation into procedures advocated by Merkhofer (1987), Morgan and Henrion (1990), Cooke (1991), and Keeney and von Winterfeldt (1991).

Spetzler and Staël Von Holstein's (1975) methods were based on what was known about the psychology of judgment in 1975. Although their methods appear to be robust to overconfidence, our understanding of the psychology of judgment has made great strides since 1975. Specifically, in addition to displaying overconfidence, subjective probability judgments do not obey extensionality and additivity principles. Different descriptions of the same event lead to different probabilities, violating the extensionality principle (e.g., Fischhoff et al. 1978, Tversky and Koehler 1994). Moreover, the judged probability of the union of disjoint events is often smaller than the sum of probabilities judged separately for these disjoint events, violating the additivity principle (e.g., Rottenstreich and Tversky 1997).

Support theory (Tversky and Koehler 1994, Rottenstreich and Tversky 1997) is a framework for subjective probability judgment that accommodates nonextensionality and subadditivity of subjective probabilities. An important feature of support theory is the support function, a modeling construct that represents how the judge summarizes the recruited evidence in favor of an event, given the way the event is described.[1] The notation $s(A)$ is used to represent the

[1] Support theory uses the term "hypothesis" to refer to the *description* of an event outcome, and the theory is developed in terms

support for event $A$; that is, it represents the individual's evaluation regarding the strength of evidence in favor of $A$. According to support theory, $s(A)$ is not *description invariant*; different descriptions of the same event do not necessarily carry the same support. For example, suppose $A$ can be decomposed into disjoint $A_1$ and $A_2$. It is possible that describing $A$ explicitly as ($A_1$ or $A_2$) can result in $s(A) \neq s(A_1 \text{ or } A_2)$. In turn, differences in support due to different descriptions can lead to different stated probabilities. In support theory, $P(A, B)$ is read as "the judged probability that event $A$ holds rather than alternative event $B$." Because only one of $A$ or $B$ can occur, $P(A, B) = s(A)/[s(A) + s(B)]$, and $P(A, B)$ thus may depend on how $A$ and $B$ are described.

Support theory typically assumes that unpacking an event $A$ into disjoint events increases total support, and the sum of the support for disjoint events is typically larger than the support of the union of those events. In symbols,

$$s(A) \leq s(A_1 \text{ or } A_2) \leq s(A_1) + s(A_2).$$

Subadditivity of support leads to the well documented subadditivity of judged probabilities (Tversky and Koehler 1994, Fox et al. 1996, Rottenstreich and Tversky 1997, Fox and Tversky 1998, Fox 1999).

Support theory is consistent with *binary complementarity* (BC), which states that $P(A, B) + P(B, A) = 1$. BC follows directly from the expression for $P(A, B)$, implicitly assuming that $P(A, B)$ and $P(B, A)$ have the same descriptions of events $A$ and $B$. Empirical evidence shows that subjective probability judgments often obey BC (e.g., Tversky and Koehler 1994, Fox and Tversky 1998), although exceptions have been documented (e.g., Brenner and Rottenstreich 1999, Macchi et al. 1999).

Support theory has provided a rich theoretical framework for studying subjective probability. However, recent research has indicated that subjective probabilities may have properties that, although not inconsistent with support theory, are not implied by it. First is a property of probability judgments known as *partition dependence* (Fox and Rottenstreich 2003, Fox and Clemen 2005, See et al. 2006), whereby an individual's subjective probability judgments about an uncertain variable can be distorted by the nature of the state space partition. For example, in asking an individual for a probability distribution for tomorrow's mid-day ambient air temperature $T$, the response may depend on how many intervals the analyst specifies for $T$ and exactly what those intervals might be. Subjective probabilities are biased toward $1/k$, where $k$ is the partition size, the number of mutually exclusive and collectively exhaustive events for which the expert assesses probabilities. This phenomenon may be interpreted as an intuitive application of the "principle of insufficient reason"; if there are no obvious reasons to consider one event more likely than others, then these events are treated as equally likely. An interpretation of the partition-dependence bias is that an individual begins with an *ignorance prior* that assigns equal probabilities to the specified events and then adjusts (usually insufficiently) to account for her beliefs.

Fox and Rottenstreich (2003) propose a model (hereafter, FR model) of subjective probabilities that combines the ignorance prior with support theory. In odds form, the FR model is given by

$$\frac{P(A, B)}{1 - P(A, B)} = \left(\frac{m_A}{m_B}\right)^{1-\lambda} \left(\frac{s(A)}{s(B)}\right)^{\lambda}, \tag{1}$$

where $m_A$ and $m_B$ are the number of elements in the partition that correspond to events $A$ and $B$, respectively, $s(A)$ and $s(B)$ represent recruited support for events $A$ and $B$, respectively, and $\lambda$ represents the weight placed on recruited support relative to the ignorance prior.

A second property that subjective probabilities may possess is *interior additivity* (IA). Originally observed by Wu and Gonzalez (1999), IA can be informally described as follows: Consider a set of disjoint events $\{A, B, C, \ldots\}$. IA is the property that $P(A \cup B) - P(B) = P(A \cup C) - P(C)$, for all events $A$, $B$, and $C$, with the conditions that neither $A$ nor $B$ nor $C$ is the null event, and that neither $A \cup B$ nor $A \cup C$ is the sure event. Thus, by "interior" we loosely mean that we restrict attention to events whose probabilities fall strictly within the unit interval. In their Experiment 2, Wu and Gonzalez (1999) observed that they could calculate the revealed or "indirect" probability of event $A$ as $P'(A) = P(A \cup B) - P(B)$ for a variety of specifications of $B$, and their various calculations of $P'(A)$ tended to be highly consistent. Furthermore, direct assessments of $P(A)$ (which we will call "direct" probabilities) tended to differ substantially from the indirect probabilities $P'(A)$.

Like partition dependence, IA is a property that is not implied by support theory. Furthermore, it is not strictly compatible with the FR model. Thus, we offer an alternative model of judged probabilities that is compatible with both IA and partition dependence. Our model, described in detail in §2, characterizes judged probability as a linear combination of ignorance prior and support components. Moreover, our model is consistent with a Bayesian interpretation of an individual's use of recruited support to update the

of hypotheses rather than events per se. The distinction between events (or more specifically *event outcomes*) and their descriptions is not necessary for our treatment; thus, for convenience we will use the term "event" throughout.

ignorance prior. Thus, our first contribution in this paper is to describe and analyze a linear model of judged probabilities, and to provide empirical evidence that supports the model, including evidence that subjective probabilities display IA. In §5.1 we will discuss in more detail how our model relates to the FR and similar models.

The second contribution of our paper concerns the question of whether support is subadditive in general as assumed in support theory. Subadditivity of support drives subadditivity of judged probabilities in support theory. In our model and in the FR model, however, subadditivity in subjective probabilities can also arise from the way in which the ignorance prior combines with support, regardless of whether recruited support is subadditive. Given the linear model, our evidence for IA suggests additivity of the support function for continuous variables.

With additive support, our model delivers a straightforward prescriptive method that analysts can use to reduce the partition-dependence bias in probability judgments for continuous variables. Our method begins by eliciting an expert's subjective probabilities using a particular framing. Following the informal discussion of IA above, we propose calculating indirect probabilities $P'(A_i)$ for a mutually exclusive and collectively exhaustive set of events $A_i$ and then normalizing these indirect probabilities to obtain $P^*(A_i) = P'(A_i)/\sum_j P'(A_j)$. According to our model, calculating $P^*$ in this way removes the effect of the ignorance prior and, hence, reduces partition dependence. Thus, the third contribution of this paper is to present this prescriptive procedure, discuss issues associated with its use, and provide preliminary empirical results comparing the performance of $P^*$ with directly assessed probabilities $P$.

The remainder of this paper is organized as follows: §2 describes our linear model; §3 provides empirical evidence for our model, including evidence showing that the support function is additive for continuous variables; and §4 presents our prescriptive procedure for counteracting partition dependence. Section 5 concludes with a general discussion of the issues that our research raises and implications for future research, both descriptive and prescriptive.

## 2. A Model of Subjective Probability Assessment

### 2.1. A Probability-Assessment Process and a Linear Probability Model

Building on Russo and Kolzow (1994), Fox and Clemen (2005), in their conclusion, describe a three-stage process that individuals go through when assessing probabilities. In this section we elaborate on Fox and Clemen's (2005) process in the context of an analyst who requests probability judgments and an expert who provides them. Our discussion provides a vehicle for developing our model of probability judgment and showing how it relates to Fox and Clemen's (2005) probability-assessment process.

**2.1.1. Stage 1: Partitioning the State Space.** In the first stage of the probability-assessment process, the expert must interpret the event for which probability judgments are required. The expert must answer two essential questions at this stage. First, what particular instances should be included in the general event under consideration? Second, how should those specific instances be categorized for the sake of making a probability judgment? For example, in the case of judging a probability distribution for the high temperature ($T$) tomorrow in New York City at 9:00 A.M., the expert would want to recall previous recorded temperatures under similar conditions as well as any reasoning that is relevant for making the judgment. The expert also has to identify appropriate categories for partitioning the state space, e.g., whether the temperature is below 0°C or above.

We assume that the analyst is able to specify the variable of interest with enough clarity that the first question is not an issue. However, with regard to the second question, we will assume that what matters is not so much what the analyst specifies in terms of categories, but what the expert subjectively adopts as a relevant partition of the state space for each judgment. Why might there be a difference? For example, an analyst might plan to assess $P(T \leq 0°C)$, $P(0°C < T \leq 5°C)$, $P(5°C < T \leq 10°C)$, and $P(T > 10°C)$. However, if at the outset the analyst asks only for $P(T \leq 0°C)$, the expert may implicitly establish a twofold partition for assessing this probability: either $T \leq 0°C$, or $T > 0°C$.

For our purposes, we will begin with the assumption that the analyst partitions the state space into $m$ elements $E_i \in \{E_1, \ldots, E_m\}$ and plans to obtain probability judgments for those $m$ elements (and possibly for events constructed from them). Continuing with the temperature example above, $m = 4$ corresponds to the four specified intervals. The event $A = \{T > 0°C\}$ is constructed of $m_A = 3$ elements. However, as described in our example above, the expert may adopt a different partition of the state space. We will let $k$ denote the number of elements in the expert's partition. Thus, if the expert adopts the twofold partition $\{T \leq 0°C, T > 0°C\}$ to assess $P(T \leq 0°C)$, the expert's subjective partition for this assessment has $k = 2$.

To formalize the discussion above, let $\Omega$ denote the state space. Let $Part_a(\Omega)$ and $Part_e(\Omega)$ represent the analyst's and expert's partitions of the state space $\Omega$, respectively. Each partition is a set of elementary events that are mutually exclusive and collectively exhaustive; from the discussion above we know that $Part_a(\Omega)$ contains $m$ elements and $Part_e(\Omega)$ contains $k$ elements. Furthermore, let $\mathcal{F}(Part_a(\Omega))$ and $\mathcal{F}(Part_e(\Omega))$ denote the corresponding sets that are generated by all possible unions of the elements, excluding the null and sure events. For example, $\mathcal{F}(Part_e(\Omega))$ contains the individual elements $E_i \in Part_e(\Omega)$, all pairwise unions, $E_i \cup E_j$, all three-way unions, and so on, up to all $(k-1)$-way unions of the $E_i$. For convenience we will let $\mathcal{F}_a$ denote $\mathcal{F}(Part_a(\Omega))$ and $\mathcal{F}_e$ denote $\mathcal{F}(Part_e(\Omega))$. $\mathcal{F}_a$ and $\mathcal{F}_e$ contain neither the null nor the sure event.

Adoption of a state space partition by the expert is consistent with support theory. For example, Brenner and Rottenstreich (1999) present evidence that judges may "repack" events that are described as disjunctions; e.g., "homicide by an acquaintance" and "homicide by a stranger" might be coalesced into a single category, "homicide." Much of the evidence for support theory comes from studies in which participants are asked to respond to probabilities one at a time, as opposed to considering the full partition at once. In these studies, judged probabilities tend to obey BC (e.g., Tversky and Koehler 1994, Fox et al. 1996, Fox and Tversky 1998), suggesting that individuals spontaneously frame an isolated probability judgment in terms of a twofold partition, the target event versus its complement.

Although repacking and spontaneously choosing a twofold partition represent situations in which $k < m$, this need not be the case. Brenner and Koehler (1999) and Dougherty and Hunter (2003) suggest that when a judgment task is framed so that the individual is able to hold all $m$ elements of the partition in her focus of attention, the judged probability of each element will reflect the support of that element relative to the sum of support for the remaining elements, i.e., $P(E_i) = s(E_i)/[s(E_1) + \cdots + s(E_m)]$. The result is that the judged probabilities sum to one.

**2.1.2. Stage 2: Evaluating Support.** The second stage in Fox and Clemen's (2005) probability judgment process is the evaluation of support for the specified event. As in support theory, we assume that the individual recruits support for the $k$ different elements in her partition by considering remembered instances, explicit reasoning, historical frequencies, and so on. However, we go beyond support theory in specifying an additive mathematical structure for the support function, for which we use notation $s^*(\cdot)$. In particular, we assume that $s^*$ has the following properties:

$$s^*(A) \geq 0 \quad \forall A \in \mathcal{F}_e, \tag{2a}$$

$$s^*(A \cup B) = s^*(A) + s^*(B) \quad \forall A, B \in \mathcal{F}_e,\ A \cap B = \phi, \tag{2b}$$

$$\sum_{i=1}^{k} s^*(E_i) = 1, \quad E_i \in Part_e(\Omega). \tag{2c}$$

The first line indicates that the individual's belief for any event $A$ is nonnegative, which is consistent with support theory's support function $s(A)$. The second and third lines further specify that $s^*$ is a fully additive function for events contained in $\mathcal{F}_e$. Thus, our $s^*$ can be loosely thought of as normalized and additive $s$, conditioned on the partition that the expert has adopted.

In this paper, we focus on judging probabilities for intervals in which a continuous uncertain quantity falls; this situation has special properties that lead us to consider additive support. First, there is often no canonical partition. Although it might be natural to think of $P(X \leq 0)$ in some cases, more generally we can think about many different ways to divide the state space. This characterization is consistent with recent research by Sloman et al. (2004) showing that support theory's assumption of subadditivity may not be universally true; with continuous variables, changing from one partition to another should not invoke more or less typical events, nor are the events more or less ambiguous. According to their arguments and results, it is not unreasonable to expect additive support for continuous variables.

**2.1.3. Stage 3: Mapping Support into State Probabilities.** The third phase in Fox and Clemen's (2005) process is the mapping of normalized support $s^*$ onto a set of numbers that represent probabilities. We model the assessed probability $P(A \mid A \in \mathcal{F}_e)$ as a linear transformation of $s^*(A)$. We begin with a general form of the model:

$$P(A \mid A \in \mathcal{F}_e) = \alpha s^*(A) + (1-\alpha)(k_A/k), \tag{3}$$

where $k_A$ denotes the number of elements that comprise event $A$, and $\alpha \in [0, 1]$ represents the weight given to $s^*(A)$. We explicitly include the condition $A \in \mathcal{F}_e$ in (3) to emphasize that the relationship holds for every event $A$ in the expert's subjective partition $Part_e(\Omega)$; recall that $A$ can be neither the null event nor the sure event. The term $(1-\alpha)(k_A/k)$ implies that probabilities will be biased away from $s^*(A)$ and toward $k_A/k$, which we will call the *ignorance prior*. Because the ignorance prior is the only term in (3) that depends on the expert's subjective partition, subjective probabilities $P$ display a partition dependence bias that arises entirely from the ignorance prior.

The model in (3) can be interpreted in a Bayesian manner. In the context of a Dirichlet model, $k_A/k$ is viewed as a prior probability that event $A$ occurs. Upon recruiting support $s^*(A)$, the expert updates the prior according to (3), where $\alpha$ is viewed as the weight given to the "new information" $s^*$. A complete and formal treatment of our Bayesian interpretation is included in the online supplement (provided in the e-companion).[2]

In contrast to typical Bayesian models, we view the ignorance prior and the resulting partition dependence bias as artifacts of the elicitation process. Rather than representing genuine prior information, the ignorance prior can be thought of as arising from a cognitive process related to memory and attention that an expert uses to assess a probability. For example, in Dougherty and Hunter's (2003) cognitive model, an expert begins by holding a set of $k$ mutually exclusive and collectively exhaustive elements of the state space in the focus of her attention, and then detects the strength of memory associated with each element. Probability is judged as the ratio of memory strength for an element relative to the sum of memory strengths for all $k$ elements. In Dougherty and Hunter's (2003) interpretation, the expert begins with zero memory strength for each element. Our model, however, suggests that the starting point is a nonzero level of memory strength that is equal for each element. The process of recruiting support leads to incremental increases in memory strength relative to the ignorance prior.

### 2.2. The Aligned Partition Model

A special version of (3) obtains when the expert adopts the same partition as the analyst, so that $Part_e(\Omega) = Part_a(\Omega)$ and, hence, $\mathcal{F}_e = \mathcal{F}_a$. Then we have

$$P(A \mid A \in \mathcal{F}_a) = \alpha s^*(A) + (1-\alpha)(m_A/m). \qquad (4)$$

This case represents the situation in which the expert is able to hold all $m$ elements of the analyst-specified partition in the focus of her attention during the assessments. Thus, we call (4) the *aligned partition model* to reflect alignment between the analyst and expert partitions. This model yields additive probabilities, as suggested by Brenner and Koehler (1999) and Dougherty and Hunter (2003).

### 2.3. The Binary Partition Model

We will be particularly interested in the case where $k = 2$. This situation is typified by asking an expert for probabilities $P(A)$ one at a time, in which case the expert may spontaneously adopt the binary partition $\{A, \sim A\}$. Thus, (3) becomes

$$P(A \mid A \in \mathcal{F}_e) = \alpha s^*(A) + (1-\alpha)/2. \qquad (5)$$

In this case, the ignorance prior term is 1/2, and it is clear that (5) obeys BC. Moreover, if $m > 2$, the judged probabilities will be subadditive. We will call (5) the *binary partition model*.

[2] An electronic companion to this paper is available as part of the online version that can be found at http://mansci.journal.informs.org/.

In summary, our linear model is a weighted combination of recruited support and an ignorance prior term that depends on the expert's partition. The presence of the ignorance prior term thus implies that subjective probabilities will display partition dependence. In the next section we develop descriptive implications, focusing on the aligned and binary partition models.

### 2.4. Descriptive Implications

A descriptive model of probability judgment should reflect known empirical properties of human judgments. For example, all three forms of our model are consistent with BC. All of the models are consistent with the idea of an ignorance prior and, hence, are also consistent with the partition dependence bias. Our model is consistent with a number of additional properties as follows (complete mathematical details included in the online supplement):

- *Subadditivity*. The aligned partition model produces additive probabilities, but the binary partition model produces subadditive probabilities if $m > 2$.
- *Interior additivity*. Both models are consistent with IA. For the aligned partition model, this is trivially true, because the model's probabilities are additive. To see this for the binary partition model, let $E_{ij} = E_i \cup E_j$ for disjoint $E_i$ and $E_j$. Then,

$$\begin{aligned} P'(E_i) &= P(E_{ij}) - P(E_j) = \alpha[s^*(E_i) + s^*(E_j)] - \alpha s^*(E_j) \\ &= \alpha[s^*(E_{il}) - s^*(E_l)], \end{aligned}$$

which is true for any events $E_i$, $E_j$, and $E_l$. For both the aligned and binary partition models, IA derives from the additivity of $s^*$. Thus, showing additivity for the aligned partition model and IA for the binary partition model will strongly suggest that $s^*$ is indeed additive.

- *Indirect vs. direct probabilities*. Additivity in the aligned partition model means that $P'(E_i) = P(E_i)$. However, because $\alpha \in [0, 1]$, in the binary partition model $P'(E_i) \leq P(E_i)$.
- *Superadditivity of indirect probabilities*. Because $\alpha \in [0, 1]$, in the binary partition model the indirect probabilities are superadditive (sum to less than one). In contrast, additivity in the aligned partition model means that indirect probabilities sum to one.

These properties are summarized in the form of testable hypotheses in Table 1, and these hypotheses will guide our empirical analysis in §3.

## 3. Empirical Evidence

### 3.1. The Aligned Partition Model: Study 1

All of the aligned partition model's predictions in Table 1 follow from the fact that the model predicts additive probabilities when the expert adopts

**Table 1 Hypotheses and Predictions from Aligned and Binary Partition Models**

| Hypothesis | Model predictions: Aligned partition model $k = m$ | Model predictions: Binary partition model $k = 2$ |
|---|---|---|
| H1 Interior additivity | H1A: $P(E_{ij}) - P(E_j) = P(E_{il}) - P(E_l)$ | |
| | H1B: $\sum_{i=1}^{m} P'(E_i)$ constant | |
| H2 Indirect probabilities vs. direct probabilities | Equality $P'(E_i) = P(E_i)$ | Less $P'(E_i) \le P(E_i)$ |
| H3 Indirect probabilities | Additive $\sum_{i=1}^{m} P'(E_i) = 1$ | Superadditive $\sum_{i=1}^{m} P'(E_i) \le 1$ |

the analyst's partition. Put another way, if the analyst can make clear the partition for which he requires probabilities, and if the expert is able to keep this partition in mind during all of the assessments, she should produce probabilities that are additive. However, those probabilities should also display partition dependence due to the ignorance prior term in (4). Our first experiment tests these predictions.

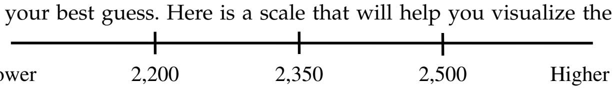
**3.1.1. Method.** Our participants in Study 1 were 145 incoming MBA students at Duke University in the fall of 2005. They completed our paper-and-pencil survey as part of an orientation session about faculty research. Each participant completed a questionnaire in which they provided their probabilities about the high temperature in Durham, NC, on a future date, and about the closing value of the NASDAQ index on a future date. Figure 1 shows an example of the NASDAQ item. To check for partition-dependence effects, we used two different partitions for each item as shown in Figure 1. Participants saw graphically how the number line was partitioned into four intervals and also were able to see all of their assessments. Thus, we made it easy for a participant to keep all four elementary events in the focus of her attention, and so we assume that the probability assessments were made with $k = m$.

Each participant saw one high partition and one low partition (temperature low/NASDAQ high partition, $N = 69$; temperature high/NASDAQ low partition, $N = 76$). Each partition had $m = 4$ elementary events (intervals), and the participants provided nine assessments for each item, 18 assessments in total. No constraints were placed on their probabilities. We counterbalanced the order of NASDAQ and temperature items as well as the order (increasing or decreasing) of the probabilities. Participants also provided knowledge ratings on a scale from 1 (I know nothing) to 7 (I know a lot).

**3.1.2. Analysis.** Our participants indicated approximately the same amount of knowledge for the

**Figure 1 Example Item and Partitions Used in Study 1**

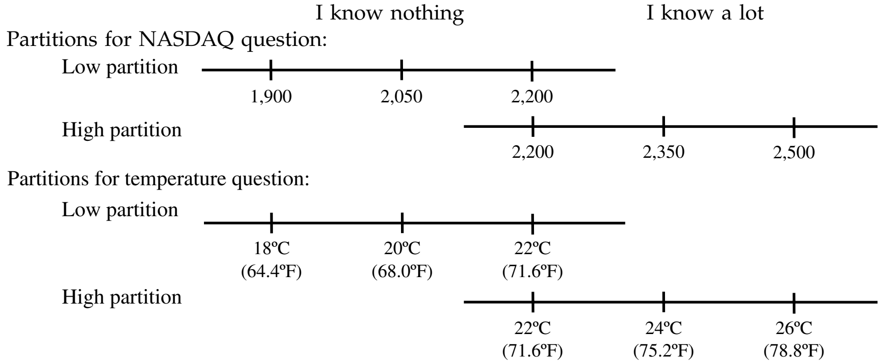
This question asks for your best estimate of the probability that the NASDAQ Stock Index will close in a designated range on January 6, 2006. If you have no idea, please give your best guess. Here is a scale that will help you visualize the intervals we will ask about:




1a. What is your best estimate of the probability (in percent) that, on January 6, 2006, the NASDAQ Stock Index will close

| | | | |
|---|---|---|---|
| Below 2,200 | ___% | Below 2,350 | ___% |
| Between 2,200 and 2,350 | ___% | Between 2,200 and 2,500 | ___% |
| Between 2,350 and 2,500 | ___% | Above 2,350 | ___% |
| Above 2,500 | ___% | Below 2,500 | ___% |
| | | Above 2,200 | ___% |

1b. Please indicate your knowledge of the NASDAQ: 1 - - - 2 - - - 3 - - - 4 - - - 5 - - - 6 - - - 7
I know nothing I know a lot




*Note.* Participants saw the partition as a number line with markers denoting intervals, and they provided their probabilities and knowledge ratings as indicated.

two questions; the average (standard deviation) of the knowledge ratings was slightly below the middle of the scale at 3.25 (1.46) for temperature, and 3.15 (1.61) for NASDAQ. We excluded temperature judgments from two participants and NASDAQ judgments from three participants because these participants did not provide all the assessments required. The remaining assessed probability distributions ($N = 143$ for temperature, $N = 142$ for NASDAQ) are analyzed below.

Of our participants, 115 assessed probabilities for the individual temperature intervals that sum to one, and 115 did the same for NASDAQ. Not surprisingly, the overlap between these two is substantial; 105 of our 143 participants provided probabilities that sum to one for both variables. To achieve full additivity, however, our participants' probabilities must satisfy five additional constraints. These constraints require that the probability for a larger interval be equal to the sum of probabilities for the smaller component intervals. For example, in the low partition for temperature, $P(\text{temp} \leq 20°\text{C}) = P(\text{temp} \leq 18°\text{C}) + P(18°\text{C} < \text{temp} \leq 20°\text{C})$; see the online supplement for full details. We investigated the extent to which the assessments display full additivity in two ways. First, for a given assessment, we calculated the number of constraints that are satisfied. Table 2 shows the results of these calculations. For both temperature and NASDAQ, 79 (more than half) of the assessments are fully additive, and 64 of our participants gave fully additive probabilities for both variables. The table also suggests that a subset of our participants had a poor understanding of probabilities or took less care with the assessments, satisfying at most one of the six constraints for each variable. Overall, the correlation between the number of constraints satisfied for temperature and for NASDAQ is 0.78 ($p < 0.0001$).

Our second approach to studying additivity in the data is to consider how much adjustment (addition or subtraction) to the assessed probabilities is needed in order to achieve full additivity; an individual's judgments that are "close" to additivity require only small adjustments, whereas those that are very far from additivity need substantially more. For each individual's assessments, we solved a linear program to find the minimum required total adjustment, i.e. the minimum sum of absolute values of all amounts added to or subtracted from the individual probabilities in order to reach full additivity. Although it is difficult to tell what constitutes a "large" or "small" total adjustment, our analysis suggests that those participants who satisfied four to five constraints required about the same amount of adjustment as would an individual whose probabilities are randomly perturbed by a normal error with standard deviation around 0.05. These are relatively small errors, suggesting that those who satisfied four to five constraints may have made simple calculation errors. Full details of this analysis are included in the online supplement.

The results from Study 1 suggest that, when the partition is made clear, many individuals naturally produce fully additive probabilities in straightforward probability judgment tasks. A still greater percentage produce probabilities that sum to one. Our participants performed under less-than-optimal conditions; our questionnaire contained minimal instructions and was one of several in a packet to be completed during a one-hour session. Thus, we take the results as evidence for the aligned partition model and for the additivity of $s^*$.

Even fully additive assessments may still exhibit partition dependence, so we tested our data for this bias. Table 3 reports the median probabilities in each partition condition. The table shows medians for the full data set, the subset of assessments that sum to one, and the subset of fully additive assessments. The

**Table 2 Additivity of Probability Assessments in Study 1**

| | NASDAQ | Temperature | Combined |
|---|---|---|---|
| *N* | 142 | 143 | 140 |
| Sum to one | 115 | 115 | 105 |
| Constraints satisfied: | | | |
| 0–1 | 34 | 31 | 22 |
| 2–3 | 14 | 18 | 5 |
| 4–5 | 15 | 15 | 3 |
| 6 (fully additive) | 79 | 79 | 64 |

*Notes.* "Sum to one" indicates the number of responses that added to one, and the following four rows indicate the number of responses that satisfied the indicated number of probability constraints. The "Combined" column indicates the number of participants who responded to both questions (*N*) and the number that fell in the same category (row) for both NASDAQ and temperature.

**Table 3 Median Probabilities in Study 1**

| | Temperature | | | NASDAQ | | |
|---|---|---|---|---|---|---|
| Partition | ≤22°C | >22°C | *n* | ≤2,200 | >2,200 | *n* |
| All data | | | | | | |
| Low | **0.42** | 0.60 | 68 | **0.90** | 0.10 | 75 |
| High | 0.10 | **0.90** | 75 | 0.20 | **0.80** | 67 |
| Sum to one | | | | | | |
| Low | **0.42** | 0.58 | 54 | **0.90** | 0.10 | 61 |
| High | 0.10 | **0.90** | 61 | 0.20 | **0.80** | 54 |
| Fully additive | | | | | | |
| Low | **0.45** | 0.55 | 39 | **0.90** | 0.10 | 40 |
| High | 0.10 | **0.90** | 40 | 0.20 | **0.80** | 39 |

*Notes.* The boldface entries are the sum of judged probabilities. For example, in the low partition condition for temperature, $P(\text{temperature} \leq 22°\text{C})$ is the sum of three judged probabilities, and the median of these sums is 41.5%. In contrast, in the high partition, $P(\text{temperature} \leq 22°\text{C})$ is a single probability. Rows "Sum to one" and "Fully additive" refer to subsets of the data set that satisfy the indicated condition.

boldface entries are median sums of judged probabilities. For example, $P(\text{Temp} \leq 22°\text{C}) = P(\text{Temp} \leq 18°\text{C}) + P(18°\text{C} < \text{Temp} \leq 20°\text{C}) + P(20°\text{C} < \text{Temp} \leq 22°\text{C})$. In the low condition, the median $P(\text{Temp} \leq 22°\text{C})$ is 42%, but in the high condition the same probability is 10%. Likewise, for the NASDAQ, in the low condition, the median $P(\text{NASDAQ} \leq 2{,}200)$ is 90%, but only 20% in the high condition, showing a strong partition dependence effect. To perform an overall test of partition dependence, we first calculated $P(\text{Temp} \leq 22°\text{C}) + P(\text{NASDAQ} > 2{,}200)$ for each participant. Because each participant saw both a high and a low partition, this sum is either a sum of two individual probabilities (for those who saw temperature high/NASDAQ low) or a sum of six probabilities, three of each variable (for those who saw temperature low/NASDAQ high). Partition dependence suggests that the sums for individuals in the temperature low/NASDAQ high condition will be greater than the sums for their counterparts in the temperature high/NASDAQ low condition. Testing this prediction with the Wilcoxon statistic, the results for each of the three subsets of the data are highly significant ($p < 0.0001$ in each case). Thus, our results extend those of Fox and Clemen (2005), who showed partition dependence when judges were required only to give probabilities for partition elements and to ensure that those probabilities summed to one.

For many decision analysts, the primary quality criterion for expert probabilities is that they be coherent. Our results in Study 1 suggest that it is not difficult to obtain additive probabilities; under less-than-ideal conditions and with no instruction to check their probabilities, approximately half of our participants provided fully additive probabilities, and many more provided probabilities that sum to one. However, although many of the probabilities were additive, the degree of partition dependence was very high across the board. Thus, the message for decision analysts is that obtaining coherent probabilities appears to be much less problematic than reducing partition dependence. Although more careful instructions and incentive-compatible compensation might improve the results (more additive responses as well as less partition dependence), the results in Fox and Clemen (2005) suggest that, even under such conditions, a high degree of partition dependence will remain.

### 3.2. The Binary Partition Model: Evidence from Tversky and Fox (1995)

Although our first experiment provides evidence for additivity and the aligned partition model, it was designed to ensure that the participants would adopt the analyst's (in this case the experimenter's) partition, so that $k = m$. To test the predictions of the binary partition model, we need a different situation, one in which the participants are not face-to-face with the experimenter's partition and are free to adopt their own. We have speculated above that, when asked to assess probabilities one at a time, individuals naturally adopt a subjective partition that is twofold, i.e., $\{A, \sim A\}$, so that $k = 2$.

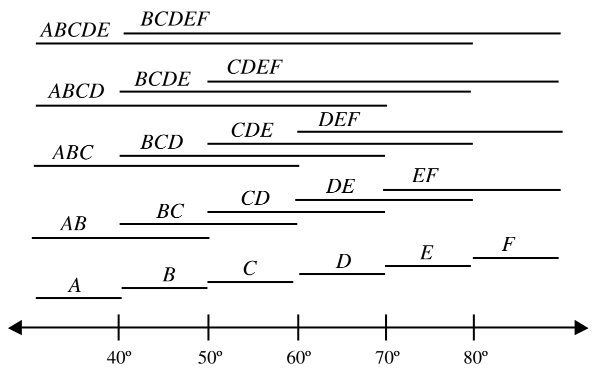
Tversky and Fox (1995) collected data (hereafter, TF data) that can be used to test the hypotheses in Table 1. Their participants made subjective probability judgments on five as-yet-unrealized events: change in the Dow Jones Industrial Average over the following week, daytime high temperatures in San Francisco and Beijing on a given future date, the score differential between the Portland Trailblazers and the Utah Jazz in the first game of the 1991 National Basketball Association (NBA) quarter final series, and the score differential between the Buffalo Bills and the Washington Redskins in the 1992 Super Bowl. Below we provide a brief description of the relevant aspects of the data. The interested reader is referred to Tversky and Fox (1995) for full details.[3]

**3.2.1. Method.** For the NBA question, the participants were 28 male Stanford students who responded to advertisements calling for basketball fans to take part in a study of decision making. The same subjects also gave probabilities for the San Francisco temperature. The participants in the Super Bowl study were 43 male football fans recruited in a similar manner; they also answered the questions on the Dow Jones closing index. Finally, 47 Stanford students in an introductory psychology course answered probability questions on San Francisco and Beijing temperature. We pooled the responses to the San Francisco temperature questions because there were no systematic differences across the two studies in which these responses were collected. Pooling gave a total sample size of 75 in this case.

The state space for the San Francisco temperature question was divided into six intervals, and 20 assessments were elicited (see Figure 2). For example, a participant judged the probability of each interval ($A, B, C, D, E$, and $F$), as well as the probability for larger contiguous intervals such as $AB$, $BC$, and so on. For each interval, the participant was asked for his or her probability for that interval versus its complement; thus, the elicitation was structured to encourage the participant to invoke $k = 2$ for each probability response. The elicitation scheme for the Beijing temperature question was similar. For the NBA question, the state space was divided into 10 intervals, and participants provided 32 probability assessments. For the Super Bowl and Dow Jones questions, the state

[3] The TF data were generously provided to us by Craig Fox.

**Figure 2** Partition and Probabilities Assessed for San Francisco and Beijing Temperatures




spaces were divided into eight events, and participants gave 28 assessments for each of the two uncertain variables.

**3.2.2. Analysis.** The first question to ask is whether our assumption that $k = 2$ for each probability is sensible. The simplest way to answer this question is to study the extent to which the probabilities displayed BC but not threefold complementarity. The results of our analysis are in Table 4. For each of the five studies, it was possible to analyze multiple partitions; the second row in Table 4 shows the number of possible partitions in each study. For each of the partitions, we used a sign test to test the hypothesis that the median sum of probabilities equals one. The remaining rows of the table show the number of tests that were significant at the indicated level.

The left-hand portion of Table 4 shows results for binary partitions. Four of the 43 tests were significant at the 0.01 level and three more at the 0.05 level. Although not shown, almost all of the 43 medians were close to 1.00; 31 of the medians equaled 1.00, eight were within 5% of 1.00, and three were within 10%. The right-hand portion of Table 4 shows results for threefold partitions. Of the 60 possible partitions, only seven had $p$-values greater than 0.05. Nine were significant at the 0.05 level, 11 more at the 0.01 level, and 33 at the 0.001 level. In fact, 58 of the 60 median sums were greater than 1.00. Thus, Table 4 provides evidence for both BC and subadditivity in the TF data. In the context of our binary partition model, we take these results as evidence that the judges indeed tended to invoke a $k = 2$ partition.

Given that the binary partition model holds, the TF data present numerous ways to test the hypotheses in Table 1. For example, for each participant who assessed probabilities for the San Francisco temperature, the indirect probability of event $A$, $P'(A)$, can be calculated in four different ways: $P(AB) - P(B)$, $P(ABC) - P(BC)$, $P(ABCD) - P(BCD)$, and $P(ABCDE) - P(BCDE)$. Hence, for event $A$ only, there are six ways to test Hypothesis H1A by comparing one version of $P'(A)$ with another. We can make similar calculations for elemental events $B$, $C$, $D$, $E$, and $F$, and for events like $AB$, $BC$, $ABC$, and so on. Because the TF probabilities were elicited one at a time, however, they exhibit inconsistencies, and approximately 14% of the calculated indirect probabilities ($P'$) are negative. (The percentage of negative $P'$ ranges from 9% in the Super Bowl data set to 17% for the San Francisco temperature data.) In all of the remaining analyses, we have set all negative $P'$ to zero. Doing so represents an informal first step to remove inconsistencies. It also results in a substantial reduction of noise in the data by removing outliers while otherwise affecting the data in relatively minor ways, as the following analysis shows.

Figures 3 and 4 give an idea of the extent of the inconsistencies in the assessed probabilities, as manifested in variability in $P'$. In Figure 3, for each data set we show the range of $P'$—calculated setting negative indirect probabilities to zero as indicated above—for a typical individual[4] in that data set. The variability is striking in some instances; e.g., for event $F$ in the San Francisco data set, $P'(F)$ ranges from zero to 0.5, with an average of 0.15. The median range, taken over all 30 events, is 0.235. The "×" symbols in the graph represent the average of $P'$ including the negative probabilities, whereas "−" represents the average of $P'$ where negative values are set equal to zero; in only eight out of 30 events is there a noticeable difference, suggesting that setting negative $P'$ equal to zero has relatively little effect.

Although the ranges shown in Figure 3 are instructive, they give no idea how these ranges would vary across individuals. To provide an idea of the variability across individuals, we constructed Figure 4. We first identified a typical event[5] in each data set. For that event we calculated the range of indirect probabilities for every individual. The box-and-whisker plots in Figure 4 represent the variability in the individuals' ranges. For each data set, the most consistent individuals have ranges at or near zero, regardless of whether negative indirect probabilities are set to zero. The medians and the interquartile ranges shown in the box-and-whisker plots indicate that in general there is some inconsistency in the indirect probabilities, and the greatest variation in ranges occurs in the

[4] To identify the typical individual in each data set, we first calculated the ranges of the indirect probabilities for all six events for each individual in the data set. We then calculated the average range for each individual. We chose as the typical individual the one with the median of the average ranges.

[5] To identify the typical event, we first calculated $P'$ for each event and for each individual, and for each individual we normalized the $P'$ to sum to one. We then found the median normalized $P'$ for each event. The typical event for each data set is the event with the median of the median normalized $P'$.

**Table 4 Complementarity in Tversky and Fox's (1995) Data: Twofold and Threefold Partitions**

| | Twofold (binary) partitions | | | | | Threefold partitions | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Beijing | San Francisco | Dow Jones | Super Bowl | NBA | Beijing | San Francisco | Dow Jones | Super Bowl | NBA |
| $N$ | 47 | 75 | 43 | 43 | 28 | 47 | 75 | 43 | 43 | 28 |
| Total number of partitions | 5 | 5 | 10 | 10 | 13 | 10 | 10 | 14 | 14 | 12 |
| $p \geq 0.20$ | 2 | 4 | 10 | 8 | 12 | — | — | 2 | 1 | — |
| $0.05 \leq p < 0.20$ | — | — | — | — | — | — | — | 2 | — | 2 |
| $0.01 \leq p < 0.05$ | 1 | 1 | — | 1 | — | — | 1 | 3 | 1 | 4 |
| $0.001 \leq p < 0.01$ | 1 | — | — | 1 | 1 | 4 | — | 5 | 1 | 1 |
| $p < 0.001$ | 1 | — | — | — | — | 6 | 9 | 2 | 11 | 5 |

*Notes.* $N$ indicates the sample size. In each study, it was possible to consider multiple twofold and threefold partitions; the total number of possible partitions is shown in the second row. For each partition we ran a sign test on the null hypothesis that the median sum of probabilities equals 1.00. The last four rows show the number of partitions for which the sign test $p$-value fell in the indicated range. The $p$-values are two-sided.

San Francisco data set. However, the primary effect of setting negative indirect probabilities equal to zero is to remove the outliers, thereby reducing noise in the data as intended, especially in the Beijing, San Francisco, and Super Bowl data sets.

Along with the noise reduction effect, we argue that setting negative indirect probabilities to zero results in no substantial bias when testing for interior additivity (H1A and H1B); the effects on the different indirect probabilities are roughly symmetric. In comparing indirect with direct probabilities (H2) and testing for superadditivity of indirect probabilities, setting negative indirect probabilities to zero results in more conservative tests. Thus, the deleterious effects of setting negatives to zero appear to be relatively minor and should be outweighed by the noise reduction achieved and the resulting increase in statistical power in our tests.[6]

Our first test of interior additivity (H1A) requires the comparison of $P'$ calculated in different ways. We used the Friedman nonparametric test to compare the different calculations of $P'$; in the typical language of the Friedman test, the different methods are the treatments and the individual participants are the blocks. We ran a total of 80 separate tests, one for each elementary event in each of the five studies. (See Table 5 for the number of tests conducted for each study.) Because our linear model predicts that $P'$ will be the same regardless of the method used to calculate it, failing to reject the null hypothesis of equality is evidence for the model. Table 5 shows the number of significant results in each data set for four different threshold $p$-values. Overall, 10 tests were significant for $p < 0.001$, seven for $p$ between 0.001 and 0.01, nine for $p$ between 0.01 and 0.05, and 15 for $p$ between 0.05 and 0.20.

[6] As a check, we also performed all calculations without changing the negative indirect probabilities. The online supplement contains versions of Tables 5–8 based on these calculations. The results are qualitatively similar to those presented here.

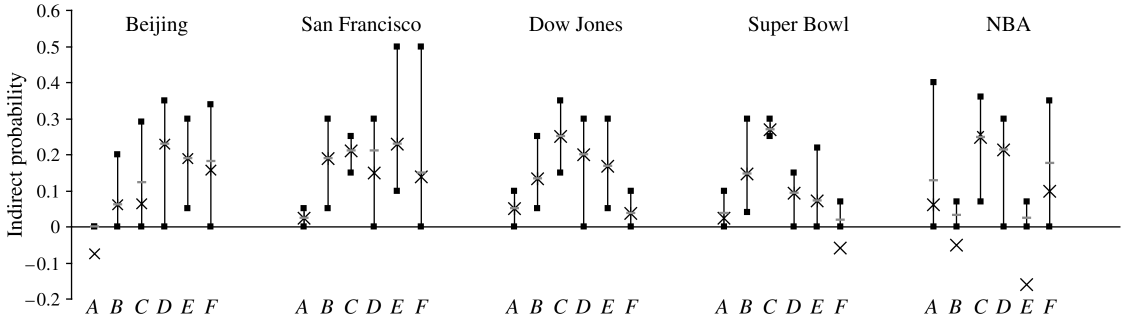
Although there are relatively few significant results among our 80 tests, these results do not provide overwhelming support for H1A, equal indirect probabilities. For example, about 51% of the results were significant at the 0.20 level or greater; for other significance levels, the proportion of significant results is also greater than one would expect if the null hypothesis of equal indirect probabilities were true.

H1B, constant sum of $P'$, provides an indirect approach to testing for IA. To test H1B, we began by finding the sum of $P'$ using three different methods. In the first method, we summed $P'$ of the form $P(E_iE_j) - P(E_j)$ over all the primary events in the partition (events $A, B, C, D, E$, and $F$); for example, $P'(A)$ in this method is calculated as $P(AB) - P(B)$.

**Table 5 Testing Hypotheses H1A, H1B, H2, and H3 for the Binary Partition Model: Summary Results**

| | Beijing | San Francisco | Dow Jones | Super Bowl | NBA |
|---|---|---|---|---|---|
| $N$ | 47 | 75 | 42 | 42 | 28 |
| H1A *Equal indirect probabilities* | | | | | |
| Test: Friedman | | | | | |
| No. of tests | 16 | 16 | 15 | 15 | 18 |
| No. of results significant at: | | | | | |
| $p < 0.001$ | 0 | 3 | 2 | 3 | 2 |
| $0.001 \leq p < 0.01$ | 0 | 1 | 2 | 2 | 2 |
| $0.01 \leq p < 0.05$ | 1 | 3 | 2 | 1 | 2 |
| $0.05 \leq p < 0.20$ | 3 | 3 | 3 | 5 | 1 |
| H1B *Constant sum of indirect probabilities* | | | | | |
| Median sum of indirect probabilities: | | | | | |
| Method 1 | 0.800 | 0.825 | — | — | — |
| Method 2 | 0.650 | 0.775 | — | — | — |
| Method 3 | 0.650 | 0.700 | — | — | — |
| Test: Friedman $p$-value | 0.360 | 0.639 | — | — | — |
| H2 *Indirect probabilities less than direct* | | | | | |
| Median sum of probabilities: | | | | | |
| Indirect | 0.758 | 0.790 | 0.958 | 0.926 | 1.005 |
| Direct | 1.900 | 1.850 | 1.250 | 1.480 | 1.375 |
| Test: Wilcoxon $p$-value | <0.001 | <0.001 | <0.001 | <0.001 | <0.001 |
| H3 *Superadditivity of indirect probabilities* | | | | | |
| Test: Sign $p$-value | <0.001 | <0.001 | 0.280 | <0.019 | 0.851 |

**Figure 3 Variability of Indirect Probabilities for a Typical Individual**




*Notes.* For each data set, the graph shows the range of indirect probabilities for a typical individual (see Footnote 4). The "−" symbol shows the average of the indirect probabilities for the corresponding event when negative indirect probabilities are set to zero. In comparison, the "×" symbol marks the average when negative indirect probabilities are included.

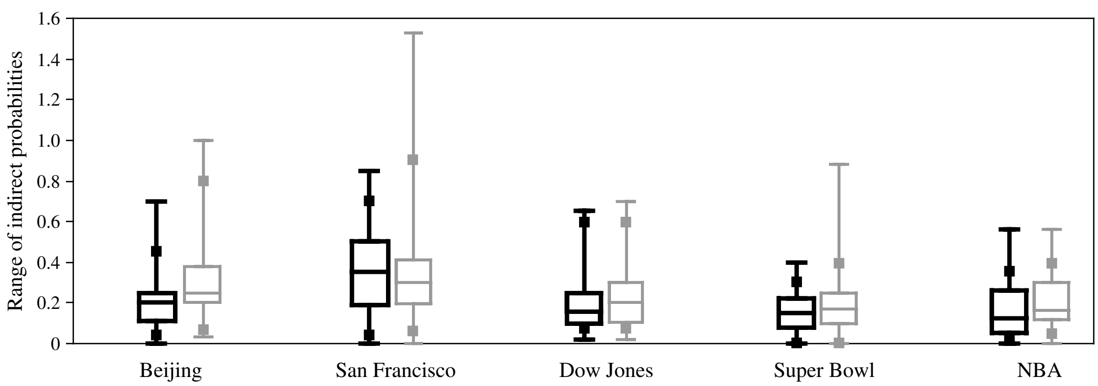
In the second method, we summed $P'$ of the form $P(E_iE_jE_k) - P(E_jE_k)$; for example, $P'(A)$ in this case is calculated as $P(ABC) - P(BC)$. In the third method, we summed $P'$ of the form $P(E_iE_jE_kE_l) - P(E_jE_kE_l)$; here $P'(A)$ is calculated as $P(ABCD) - P(BCD)$. We were not able to use indirect probabilities calculated using "larger" events, such as $P(ABCDE) - P(BCDE)$, because there were not enough assessments to make such calculations.

Completing the test of H1B, we compared the three different sums for each individual using Friedman's test. The null hypothesis is that IA holds (i.e., that the three sums of indirect probabilities are equal); thus, failing to reject the null provides support for our model. The results are shown in the second panel of Table 5. Both the San Francisco and Beijing temperature data sets were insignificant as expected ($p = 0.360$ for Beijing; $p = 0.639$ for San Francisco). Because of the particular partitions used in the Dow Jones, Super Bowl, and NBA studies, we were unable to calculate all the indirect probabilities required, so we could not perform the test on these data sets. Thus, our results for H1B provide some additional support for IA.

To test whether the indirect probabilities are less than the direct probabilities (H2), we began by calculating $\bar{P}'$, the average indirect probability for each primary event in the partition. This average was calculated over all possible ways to find $P'$, setting negatives to zero as explained above. We then calculated the sum of $\bar{P}'$ for each individual as well as the sum of the direct probabilities. If our hypothesis is true, the sum of $\bar{P}'$ is less than the sum of the direct probabilities. As shown in the third panel in Table 5, we reject the null hypothesis of equality ($p < 0.001$) in all five data sets.

Our final hypothesis is that the indirect probabilities are superadditive (H3). To test this, we used the sign test to compare the sum of the $\bar{P}'$ (as calculated above in the test of H2) with 1.00. The bottom panel in Table 5 shows that we were able to reject the null hypothesis that the median sum equals 1.00 for three of the data sets ($p < 0.001$ for Beijing and San Francisco; $p = 0.019$ for Superbowl), but not the other two ($p = 0.280$ for Dow Jones; $p = 0.851$ for NBA). As mentioned above, setting negative $P' = 0$ gives conservative results; when we

**Figure 4 Ranges of Indirect Probabilities**




*Notes.* For a typical event in each data set (see Footnote 5), the box-and-whisker plots show the variability in the ranges of indirect probabilities across individuals. The darker (left-hand) plots represent the case when negative indirect probabilities are set to zero, whereas the lighter (right-hand) plots represent the case when negative indirect probabilities are included.

included the original negative indirect probabilities, the results were $p < 0.001$ for Beijing, San Francisco, and Superbowl; $p = 0.003$ for Dow Jones; and $p = 0.172$ for NBA.

### 3.3. Empirical Evidence for the Linear Model: Summary

Our empirical results generally support the linear model. In particular, the results of Study 1 support the aligned partition model when the expert can keep all elements of the partition in the focus of her attention throughout the elicitation. Similarly, our analysis of the TF data provides support for the binary partition model, including IA as a property of indirect probabilities and, implicitly, the additivity of $s^*$.

## 4. A Prescriptive Procedure

### 4.1. Using the Linear Model

Our model and arguments above can be interpreted as indicating that partition dependence does not necessarily imply irrationality; the expert begins with a prior and appropriately updates it based on information. Partition dependence does, however, present a problem for the decision analyst, who would prefer to elicit the expert's $s^*$ without the contamination of the ignorance prior. We can use our model as the basis for developing a procedure to accomplish this goal.

Our strategy is to use the binary partition model. Calculating indirect probabilities leads to a cancellation of the common ignorance prior term, with the result that the indirect probability $P'(A) = \alpha s^*(A)$ is proportional to $s^*(A)$ (see Equation (EC.6) in the online supplement). Thus, we suggest the following:

1. Elicit probabilities for intervals in isolation so that the binary partition model ($k = 2$) holds.
2. Elicit probabilities for enough overlapping intervals so that indirect probabilities $P'$ can be calculated for all of the intervals of interest.
3. Calculate and normalize the indirect probabilities to recover $s^*$:

$$P^*(E_i) = \frac{P'(E_i)}{\sum_{j=1}^{m} P'(E_j)} = \frac{\alpha s^*(E_i)}{\alpha \sum_{j=1}^{m} s^*(E_j)} = s^*(E_i).$$

The resulting $P^*$ represents the expert's recruited evidence $s^*$ untainted by the ignorance prior.

Calculating $P'$ and $P^*$ in practice will require some care. As we have done in our analyses above, we suggest that the analyst obtain probabilities of as many intervals as possible in order to calculate $P'$ in several ways. To reduce the effect of inconsistent judgments, the average ($\bar{P}'$) or the median can be calculated, which can in turn be normalized to obtain $P^*$.

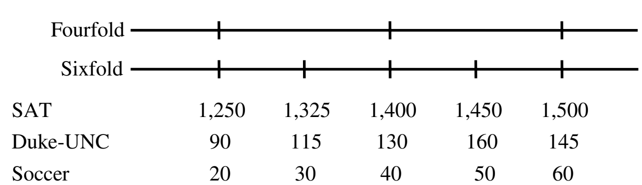
Note that the procedure requires that the binary partition be maintained for all assessments. Thus, the expert might be invited to reconsider certain assessments, provided the binary partition can be maintained. However, in the context of our linear model, reconciliation procedures that require an expert to consider the entire state space are off limits. In fact, even revisiting individual probability judgments may lead the expert away from the binary framing and toward the analyst's $m$-fold partition. We will return to the issue of practical implementation in §5.

### 4.2. $P^*$ Reduces Partition Dependence: Study 2

Although our prescriptive procedure is straightforward, a reasonable question is whether this procedure actually reduces partition dependence. The study we describe here shows that $P^*$ does tend to display less partition dependence than direct probabilities. In addition, the results indicate that $\alpha$ is related to knowledge; as suggested by the Bayesian interpretation of our model, more knowledge is associated with larger $\alpha$ and, hence, less weight on the ignorance prior term.

**4.2.1. Method.** Participants in this study were 423 Duke undergraduates in the fall of 2005. The study was conducted on laptop computers, and participants provided probability judgments for three variables: the SAT score of a randomly chosen Duke undergraduate, the total points scored by both teams in the next Duke versus University of North Carolina basketball game, and the total points (including penalty kicks) scored by all teams in the next Atlantic Coast Conference men's soccer tournament. The focus of this study is comparing direct and indirect probabilities in two different partitions, and so the partitions were set up so that the tail events were the same in the fourfold and sixfold partitions. Figure 5 shows the partitions we used for the three different questions. Participants judged probabilities in both "$k = m$" or "aligned partition" tasks in which probabilities for all

**Figure 5** **Partitions Used in Study 2**




*Notes.* Partitions were designed so that the tail events in both fourfold and sixfold partitions are the same. For example, in the fourfold Soccer partition, the intervals were $P(X \leq 20)$, $P(20 < X \leq 40)$, $P(40 < X \leq 60)$, and $P(X > 60)$. The sixfold Soccer partition had the same tail intervals along with $P(20 < X \leq 30)$, $P(30 < X \leq 40)$, $P(40 < X \leq 50)$, and $P(50 < X \leq 60)$.

individual intervals were assessed on the same computer screen, and "$k = 2$" or "binary partition" tasks in which each probability was assessed in isolation and in random order. Because the sixfold binary partition tasks involved many more probabilities to assess than the fourfold tasks, the conditions were set up so that each participant had roughly the same number of probabilities to assess; in some cases, we used filler questions to balance the assessment burden. As in Study 1, we also collected knowledge ratings for each of the three variables on a scale from 1 to 7. Four participants provided unusable data and were dropped from the analysis.

**4.2.2. Analysis.** We began by calculating $P'$ for each elementary interval in as many ways as possible, given the partitions we used and the probabilities assessed. Then, for each interval, we calculated $\bar{P}'$, setting any negative $P'$ to zero as above. Negative $P'$ accounted for 19% (23%) of all of the calculated $P'$ in the fourfold (sixfold) condition. If all of a participant's indirect probabilities for a particular interval were negative, the result was a zero probability for that interval. As above, setting negative $P' = 0$ results in substantial noise reduction and should result in no bias in comparing partition dependence in the fourfold and sixfold partitions; any effect on $P^*$ should be roughly the same for each partition.

Table 6 shows summary statistics for the three variables in Study 2. Looking at the results for the aligned partition tasks, the first observation is that most (81% overall) of the direct assessments sum to one, as in Study 1. The second observation relates to the binary partition tasks. Of the direct probability assessments, 77% overall are subadditive, whereas most (71% overall) of the indirect probabilities are superadditive. Finally, the average knowledge ratings (pooled across all participants who provided a knowledge rating for the variable, regardless of condition) show that participants believed they knew less about the Soccer question than the other two ($p < 0.0001$).

Our goal in this study is to determine whether using $P^*$ reduces partition dependence compared to the use of direct probabilities obtained from an aligned partition assessment task, which is the typical decision analysis method. We began by normalizing those direct probabilities in the aligned partition tasks that did not sum to one (again, reflecting the typical decision analysis practice of reconciling incoherent probabilities). Because the tail events were the same in the two partitions, we compared the tail probabilities (the sum of the probabilities assigned to the lower and upper tails) across the partitions. Our theory suggests that we should observe partition dependence when comparing direct tail probabilities across the two partitions in the aligned partition tasks, but not when comparing $P^*$ in the binary partition tasks. We used the Wilcoxon statistic to test this prediction, and we report one-sided $p$-values because of the directional nature of the hypothesis.

The results are shown in Table 7. In the aligned partition tasks, the median tail probability in the fourfold partition is greater than in the sixfold partition for all three variables. The difference is highly significant for SAT and Soccer (one-sided Wilcoxon $p < 0.0001$ and $p = 0.0004$, respectively) and marginally so for Duke-UNC ($p = 0.0937$). In contrast, in the binary partition tasks, none of the median $P^*$ is significantly greater in the fourfold than in the sixfold partitions. (In fact, for SAT, fourfold median $P^*$ is considerably less than sixfold median $P^*$.) Taken together, these results suggest that using normalized indirect probabilities $P^*$ does reduce partition dependence. Moreover, these data provide further evidence for interior additivity of indirect probabilities and additional support for our linear model.

**Table 6 Summary Statistics for Study 2**

| | Aligned partition, $k = m$ | | Binary partition, $k = 2$ | | | Knowledge ratings | |
|---|---|---|---|---|---|---|---|
| | | Median sum | | Median sum | | | |
| | $N$ | Direct (% = 1.00) | $N$ | Direct (% > 1.00) | Indirect (% < 1.00) | $N$ | Avg. (std. dev.) |
| SAT | | | | | | | |
| Fourfold | 56 | 1.00 (91%) | 52 | 1.21 (75%) | 0.83 (87%) | 217 | 3.29 (1.42) |
| Sixfold | 54 | 1.00 (87%) | 55 | 1.20 (67%) | 1.01 (47%) | | |
| Duke-UNC | | | | | | | |
| Fourfold | 64 | 1.00 (84%) | 79 | 1.40 (80%) | 0.85 (78%) | 283 | 3.60 (1.65) |
| Sixfold | 68 | 1.00 (69%) | 71 | 1.70 (87%) | 0.95 (59%) | | |
| Soccer | | | | | | | |
| Fourfold | 55 | 1.00 (85%) | 53 | 1.20 (66%) | 0.80 (88%)* | 218 | 1.94 (1.24) |
| Sixfold | 54 | 1.00 (69%) | 56 | 1.43 (79%) | 0.91 (66%) | | |

*$N = 52$ for Soccer, fourfold, binary partition, indirect. One subject gave usable direct probabilities, but unusable indirect probabilities.

**Table 7 Partition Dependence in Study 2**

| | Aligned partition, $k = m$ | | Binary partition, $k = 2$ | |
|---|---|---|---|---|
| | Normalized direct tail probability | | Normalized indirect tail probability ($P^*$) | |
| | $N$ | Median | $N$ | Median |
| SAT | | | | |
| Fourfold | 56 | 0.30 | 52 | 0.22 |
| Sixfold | 54 | 0.20 | 55 | 0.31 |
| Wilcoxon $p$ | | <0.0001 | | 0.990 |
| Duke-UNC | | | | |
| Fourfold | 64 | 0.30 | 79 | 0.28 |
| Sixfold | 68 | 0.20 | 71 | 0.28 |
| Wilcoxon $p$ | | 0.094 | | 0.501 |
| Soccer | | | | |
| Fourfold | 55 | 0.35 | 52 | 0.34 |
| Sixfold | 54 | 0.20 | 56 | 0.30 |
| Wilcoxon $p$ | | 0.0001 | | 0.677 |

*Notes.* The analysis was done on the tail probabilities for each variable and each partition. Wilcoxon $p$-values are one-tailed, reflecting the directional nature of the hypothesis; partition dependence suggests that the median sum of tail probabilities will be greater in the fourfold partitions than in the sixfold partitions.

Finally, data from the binary partition tasks can be used to test whether a relationship exists between weight $\alpha$ and knowledge. For each assessment in the binary partition tasks, we calculated the sum of the average indirect probabilities, $\sum \bar{P}'$, to get an estimate of $\alpha$ for that assessment (see Equation (EC.6) in the online supplement). Using data from all three variables, we regressed estimated $\alpha$ on knowledge rating, including dummy variables to indicate the SAT and Soccer variables. The results are shown in Table 8. The coefficient for knowledge rating is positive and significant ($p = 0.047$), indicating a relationship between $\alpha$ and knowledge. The coefficients for Soccer and SAT are not significant ($p = 0.279$ and 0.179, respectively), suggesting that the relationship between $\alpha$ and knowledge does not depend on the event assessed.

These results suggest that larger values of $\alpha$ are associated with higher knowledge ratings. On one hand, this is good news for decision analysts; if experts have substantial knowledge, the expert's judgments should suffer less from partition dependence. On the other hand, even an expert with a high degree of knowledge may display some partition dependence. From our results, we can estimate $E(\alpha \mid \text{knowledge} = 7)$ is 0.98 for SAT, 0.94 for Duke-UNC, and 0.91 for Soccer; even the most knowledgeable participants in our sample may be expected to display some partition dependence.

**Table 8 Study 2 Regression Results: $\alpha$ vs. Knowledge Rating**

| | | | | |
|---|---|---|---|---|
| $N$ | 365 | $F$ | 4.686 | |
| Standard error | 0.236 | $p$-value | 0.003 | |
| $R$-squared | 0.194 | | | |
| | Coefficients | Standard error | $t$ ($df = 360$) | $p$ |
| Intercept (Duke-UNC) | 0.823 | 0.035 | 23.44 | <0.0001 |
| Knowledge | 0.017 | 0.009 | 2.00 | 0.047 |
| Soccer | −0.035 | 0.033 | −1.08 | 0.279 |
| SAT | 0.040 | 0.030 | 1.35 | 0.179 |

## 5. Discussion

Our analysis and empirical results provide evidence that subjective probabilities for continuous variables conform to IA and that the linear model provides a reasonable representation of how judged probabilities relate to underlying support $s^*$. Results in §4, although preliminary, suggest that using normalized indirect probabilities $P^*$ can help reduce the partition dependence bias. In addition, our results demonstrate a knowledge effect. As we surmised, $\alpha$ increases as knowledge increases. However, our results suggest that even at high levels of knowledge, some degree of partition dependence remains.

As always, research is needed to replicate and confirm our results. In particular, in light of our results on the variability of $P'$ (Figures 3 and 4) and our direct test of IA (H1A), additional work to confirm IA as a property of $P'$ would be most helpful. Further research to test the effectiveness of $P^*$ is also called for; we will return to this topic below.

### 5.1. Relationship with Other Models

Our linear model is related to several other models that have been proposed in the literature. Here we examine the relationship of our model with that of Fox and Rottenstreich (2003) and an error model proposed by Bearden et al. (2007).

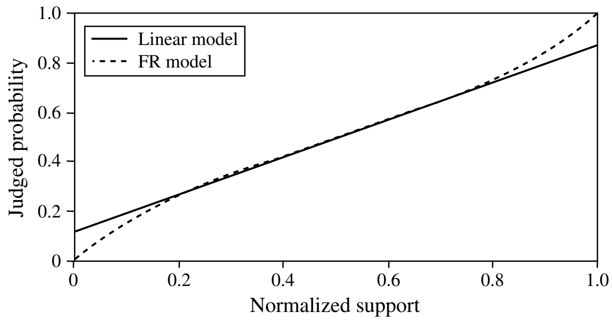
Recalling (1), the FR model is given by

$$\frac{P(A,B)}{1-P(A,B)} = \left(\frac{m_A}{m_B}\right)^{1-\lambda} \left(\frac{s(A)}{s(B)}\right)^{\lambda}.$$

Note that $P(A,B)$ is not linear in support $s(A)$. However, it is approximately linear over an interval that is strictly between 0 and 1. Figure 6 clarifies the relationship between the FR and binary partition models; our model may be thought of as exploiting the approximate linearity of the FR model within the interior of the unit interval. Empirical estimation of parameters for the FR model would help determine the curvature near the endpoints, and thus could provide a clearer idea of the range over which the linear approximation is appropriate.

Various researchers have proposed error models in which an expert's "true" or "covert" probability (like our $s^*$) is perturbed by a random error $\varepsilon$ in the process of generating the corresponding stated subjective probability $P$ (e.g., Ravinder et al. 1988, Erev

**Figure 6** Approximating the FR Model with the Binary Partition Model




*Note.* The ignorance prior term is 0.5, and $\lambda = \alpha = 0.75$.

et al. 1994, Juslin et al. 1997). In particular, Bearden et al. (2007) suggest that the pattern of partition dependence that has been observed empirically can be explained by such an error model. They assume that an individual's covert probabilities are additive but become distorted when the individual states them overtly. In their model, the distortion arises from an error process in which the stated probabilities are biased toward 1/2.

Our binary partition model is largely consistent with Bearden et al. (2007). More generally, though, our model predicts a regression effect toward $1/k$, whereas theirs predicts a regression effect toward 1/2. Thus, the real comparison between the two models may be in the context of the aligned partition model. If an expert holds all elements of the state space in the focus of her attention, are probabilities biased toward 1/2 or $1/k$? Our results, as well as those from Fox and Clemen (2005) and See et al. (2006) suggest the latter.

Regardless of the specific merits of the two models, it is interesting to note that the same prescriptive advice works in either case. Because the model by Bearden et al. (2007) predicts a consistent bias toward 1/2, calculating indirect probabilities by subtracting one direct probability from another will tend to result in offsetting errors. Normalized indirect probabilities should, therefore, display less partition dependence.

### 5.2. Future Research

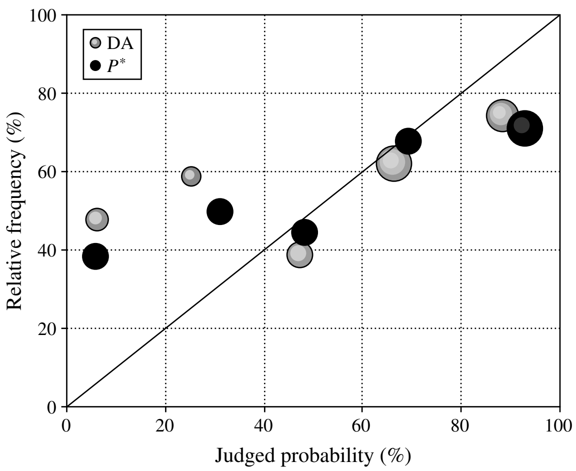
**5.2.1. Spontaneous Partition Switching and Reconciliation Processes.** Using $P^*$ successfully requires that an expert adopt a binary partition for each probability judgment and maintain this framing throughout the assessment process. However, one can imagine scenarios in which the expert spontaneously converts to an aligned partition. For example, the expert may become aware of the elementary intervals on which the elicitation is based, or the expert may find it easier to think in terms of a specific partition and use her (additive) probabilities for that partition as a basis for responding to the analyst's binary questions.

As suggested in §4.1, asking an expert to reconcile binary partition probabilities could lead the expert to switch to an aligned partition. More fundamentally, though, the process that the expert uses to reconcile the probabilities may not conform to our linear model. Understanding the psychology of the reconciliation process will determine whether and how to reconcile binary partition probabilities. For example, if the expert simply normalizes the assessed probabilities for the elementary intervals and uses those normalized probabilities to calculate the probabilities for events composed of multiple intervals, then the effect is the same as using an aligned partition, and $P^*$ will not remove the ignorance prior effect. However, experts may spontaneously use other reconciliation strategies (and it may be possible to encourage them to use such strategies) that can ultimately result in a reduction or elimination of partition dependence.

**5.2.2. Performance of $P^*$.** We showed that using $P^*$ can lead to a reduction in the partition dependence effect. However, from a decision analysis perspective, the bottom line is whether $P^*$ performs better in the sense of being better calibrated or, better still, leading to better average scores based on strictly proper scoring rules (Savage 1971, Schervish 1989, Winkler 1996). Such a study should include careful instructions and incentive-compatible compensation for a number of individuals assessing direct and indirect probabilities for a large number of different variables. Judgments would be for variables for which realizations would become known, thereby permitting the calculation of calibration statistics, average scores, and related performance measures. In a preliminary study (full details in the online supplement), we randomly assigned 85 participants to direct assessment (DA) and indirect assessment ($P^*$) conditions. Each participant gave assessments and knowledge ratings for four uncertain variables drawn randomly from a pool of 100 uncertain variables. Calibration data for both conditions are shown in Figure 7. The results suggest that $P^*$ may be better calibrated than DA, although the difference is not significant. We also calculated average scores using a quadratic scoring rule for which a higher score indicates better performance. The median score for $P^*$ was 0.262, compared to a median score of 0.250 for DA; again, although the results are not significant, they suggest that $P^*$ may perform better than DA. These results clearly are not conclusive; more comprehensive studies are under way.

Another aspect of performance evaluation would be to compare the performance of $P^*$ with a more conventional debiasing procedure. For example, Fox and Clemen (2005) suggest that the analyst work

**Figure 7 Calibration Data for Preliminary Study of $P^*$ Performance**




*Notes.* Participants' judged probabilities were grouped into five equally spaced bins. The size of the bubble reflects the number of judgments in the corresponding bin.

with an expert in advance to establish multiple partitions. Having an expert make judgments over multiple partitions may highlight inconsistencies due to partition dependence. The analyst can then work with the expert to reconcile her judgments. Whether such a procedure would perform better than $P^*$ in terms of reduced partition dependence, improved calibration, or better average scores is an empirical question.

### 5.3. When Should Analysts Use $P^*$ vs. Conventional Probability Assessment?

Using $P^*$ amounts to adjusting an expert's judged probabilities ex post, whereas conventional debiasing methods focus on changes in the elicitation process itself to counteract biases (Fischhoff 1982, Morgan and Henrion 1990, McClelland and Bolger 1994). Although ex post adjustment is not new (e.g., Lindley et al. 1979, Osherson et al. 1997), decision theorists have typically eschewed ex-post procedures. Savage (1971, p. 796) argues eloquently, "You might discover with experience that your expert is optimistic or pessimistic in some respect and therefore temper his judgments. Should he suspect you of this, however, you and he may well be on the escalator to perdition." We surmise, however, that an expert may have difficulty figuring out how to temper her judgments in anticipation of adjustments for partition dependence.

Another argument against ex post adjustment is that experts may be required to "sign off" on their judgments. If the analyst has adjusted the probabilities ex post, then it may not be clear whose beliefs $P^*$ represents. However, we offer two counterarguments. First, as suggested above, in some cases the analyst may be able to show the calculated $P^*$ to the expert and obtain confirmation that they are reasonable representations of the expert's beliefs. Second, even without such confirmation, it is not unreasonable for the expert to sign off on her original judgments and for the analyst to sign off on $P^*$ as representing the appropriate analysis and adjustment of those expert judgments. Such a procedure is already implicitly used, for example, when an analyst presents a combination of expert probabilities (e.g., Cooke 1991), certifying that appropriate weights have been applied to the expert judgments in calculating the combination.

Our arguments above notwithstanding, we hesitate to recommend using $P^*$ in all elicitations. We speculate that conventional probability assessment methods, along the lines suggested by Fox and Clemen (2005), will be most appropriate with truly knowledgeable experts and when adequate time and resources are available. However, when experts have less knowledge (and hence the effect of the ignorance prior is more pronounced) or the analyst has less time to guide the expert, using $P^*$ should help reduce partition dependence. For example, in constructing a Bayes net or influence diagram using subjective probabilities, an expert may be required to assess hundreds of probabilities (e.g., Heckerman et al. 1992, Abramson et al. 1996). Other applications may include risk analyses in which project constraints permit only initial guidance of the expert by the analyst, after which the expert must assess the required probabilities on her own. As the use of subjective probabilities becomes more widespread, we suspect that such situations will become more commonplace, and that the use of ex post procedures such as $P^*$ will in turn become more valuable.

## 6. Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at http://mansci.journal.informs.org/.

**Acknowledgments**
The authors thank Craig Fox, Yuval Rottenstreich, George Wu, two anonymous referees, and the associate editor for helpful discussions. Craig Fox in particular was the one who originally pointed them toward Wu and Gonzalez's idea of interior additivity, later suggested that they consider a linear probability model, and graciously provided the TF data. This work was supported by the National Science Foundation under Grant SES-0317867. Any opinions, findings, conclusions, or recommendations expressed herein are those of the authors and do not necessarily reflect the views of the National Science Foundation.

### References

Abramson, B., J. Brown, W. Edwards, A. H. Murphy, R. L. Winkler. 1996. Hailfinder: A Bayesian system for forecasting severe weather. *Internat. J. Forecasting* **12** 57–71.

Bearden, J. N., T. S. Wallsten, C. R. Fox. 2007. Contrasting stochastic and support theory accounts of subadditivity. *J. Math. Psych.* **51** 229–241.

Brenner, L. A., D. J. Koehler. 1999. Subjective probability of disjunctive hypotheses: Local-weight models for decomposition of evidential support. *Cognitive Psych.* **38** 16–47.

Brenner, L. A., Y. Rottenstreich. 1999. Focus, repacking, and the judgment of grouped hypotheses. *J. Behavioral Decision Making* **12** 141–148.

Cooke, R. M. 1991. *Experts in Uncertainty*: *Opinions and Subjective Probability in Science*. Oxford University Press, New York.

Dougherty, M. R. P., J. Hunter. 2003. Probability judgment and subadditivity: The role of working memory capacity and constraining retrieval. *Memory and Cognition* **31** 968–982.

Erev, I., T. S. Wallsten, D. V. Budescu. 1994. Simultaneous over- and underconfidence: The role of error in judgment processes. *Psych. Rev.* **98** 506–528.

Fischhoff, B. 1982. Debiasing. D. Kahneman, P. Slovic, A. Tversky, eds. *Judgment under Uncertainty*: *Heuristics and Biases*. Cambridge University Press, Cambridge, UK, 422–444.

Fischhoff, B., P. Slovic, S. Lichtenstein. 1978. Fault trees: Sensitivity of estimated failure probabilities to problem representation. *J. Experiment. Psych.*: *Human Perception Performance* **4** 330–344.

Fox, C. R. 1999. Strength of evidence, judged probability, and choice under uncertainty. *Cognitive Psych.* **38** 167–189.

Fox, C. R., R. T. Clemen. 2005. Subjective probability assessment in decision analysis: Partition dependence and bias toward the ignorance prior. *Management Sci.* **51** 1417–1432.

Fox, C. R., Y. Rottenstreich. 2003. Partition priming in judgment under uncertainty. *Psych. Sci.* **14** 195–200.

Fox, C. R., A. Tversky. 1998. A belief-based account of decision under uncertainty. *Management Sci.* **44** 879–895.

Fox, C. R., B. A. Rogers, A. Tversky. 1996. Options traders exhibit subadditive decision weights. *J. Risk Uncertainty* **1** 5–17.

Gilovich, T., D. Griffin, D. Kahneman, eds. 2002. *Heuristics and Biases*: *The Psychology of Intuitive Judgment*. Cambridge University Press, Cambridge, UK.

Heckerman, D., E. Horvitz, B. Nathwani. 1992. Toward normative expert systems: Part I. The Pathfinder project. *Methods Inform. Medicine* **31** 90–105.

Juslin, P., H. Olsson, M. Björkman. 1997. Brunswikian and Thurstonian origins of bias in probability assessment: On the interpretation of stochastic components of judgment. *J. Behavioral Decision Making* **10** 189–209.

Kahneman, D., P. Slovic, A. Tversky, eds. 1982. *Judgment under Uncertainty*: *Heuristics and Biases*. Cambridge University Press, Cambridge, UK.

Keeney, R., D. von Winterfeldt. 1991. Eliciting probabilities from experts in complex technical problems. *IEEE Trans. Engrg. Management* **38** 191–201.

Lindley, D., A. Tversky, R. Brown. 1979. On the reconciliation of probability assessments. *J. Roy. Statist. Soc. Ser. A* **142**(Part 2) 146–180.

Macchi, L., D. Osherson, D. H. Krantz. 1999. A note on superadditive probability judgment. *Psych. Rev.* **106** 210–214.

McClelland, A. G. R., F. Bolger. 1994. The calibration of subjective probabilities: Theories and models 1980–94. G. Wright, P. Ayton, eds. *Subjective Probability*. Wiley, New York, 453–482.

Merkhofer, M. W. 1987. Quantifying judgmental uncertainty: Methodology, experiences, and insights. *IEEE Trans. Systems, Man, Cybernetics* **17** 741–752.

Morgan, M., M. Henrion. 1990. *Uncertainty*: *A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Management*. Cambridge University Press, Cambridge, UK.

Osherson, D., E. Shafir, D. H. Krantz, E. E. Smith. 1997. Probability bootstrapping: Improving prediction by fitting extensional models to knowledgeable but incoherent probability judgments. *Organ. Behav. Human Decision Processes* **69** 1–8.

Ravinder, H. V., D. N. Kleinmuntz, J. S. Dyer. 1988. The reliability of subjective probabilities obtained through decomposition. *Management Sci.* **34** 186–199.

Rottenstreich, Y., A. Tversky. 1997. Unpacking, repacking, and anchoring: Advances in support theory. *Psych. Rev.* **104** 406–415.

Russo, J. E., K. J. Kolzow. 1994. Where is the fault in fault trees? *J. Experiment. Psych.: Human Perception Performance* **20** 17–32.

Savage, L. J. 1971. The elicitation of personal probabilities and expectations. *J. Amer. Statist. Assoc.* **66** 783–801.

Schervish, M. J. 1989. A general method for comparing probability assessors. *Ann. Statist.* **17** 1856–1879.

See, K. E., C. R. Fox, Y. Rottenstreich. 2006. Between ignorance and truth: Partition dependence and learning in judgment under uncertainty. *J. Experiment. Psych.*: *Learn., Memory, Cognition* **32** 1385–1402.

Sloman, S., Y. Rottenstreich, E. Wisniewski, C. Hadjichristidis, C. R. Fox. 2004. Typical versus atypical unpacking and superadditive probability judgment. *J. Experiment. Psych.*: *Learn., Memory, Cognition* **30** 573–582.

Spetzler, C. S., C.-A. Staël Von Holstein. 1975. Probability encoding in decision analysis. *Management Sci.* **22** 340–352.

Tversky, A., C. R. Fox. 1995. Weighing risk and uncertainty. *Psych. Rev.* **102** 269–283.

Tversky, A., D. J. Koehler. 1994. Support theory: A nonextensional representation of subjective probability. *Psych. Rev.* **101** 547–567.

Winkler, R. L. 1996. Scoring rules and the evaluation of probabilities. *Test* **5** 1–26.

Wu, G., R. Gonzalez. 1999. Nonlinear decision weights in choice under uncertainty. *Management Sci.* **45** 74–85.