Partition Dependence and Subjective Probabilities: Problem and Solution

Bob Clemen

Duke University

With Craig Fox, Canan Ulu

Psychology and Decision Analysis: The New Wave

- Using psychological models to develop precise prescriptive methods
  - Preferences
    - Peter Wakker and colleagues:
      Distortions due to Prospect Theory
    - Anderson & Hobbs: Scale compatibility bias
    - Delquie: Optimal trade-off assessment
  - Probabilities
    - Clemen & colleagues:
      Adjusting probabilities for overconfidence

→***Counteracting partition dependence***

FTSE 100 Stock Index
Closing value, Oct 21, 2005

Group 1

Assess probabilities
for these events:

- $\text{FTSE} \leq 4200$
- $4200 < \text{FTSE} \leq 4800$
- $4800 < \text{FTSE} \leq 5500$
- $5500 < \text{FTSE} \leq 6300$
- $6300 < \text{FTSE}$

Group 2

Assess probabilities
for these events:

- $\text{FTSE} \leq 2800$
- $2800 < \text{FTSE} \leq 3200$
- $3200 < \text{FTSE} \leq 3700$
- $3700 < \text{FTSE} \leq 4200$
- $4200 \leq \text{FTSE}$

Overview

- The partition dependence phenomenon
- A psychological model of probability judgment
- A prescriptive solution

Beyond Heuristics and Biases: **Support Theory**

- Given a specified event partition:

A | ~A

$$P_A = \frac{S_A}{S_A + S_{\sim A}}$$

Tversky and Koehler (1994)
Rottenstreich & Tversky (1997)

A | B | C

$$P_A = \frac{S_A}{S_A + S_B + S_C}$$

Brenner & Koehler (1999)

A is called the "focal" hypothesis; ~A or B∪C is "alternative."

S is positive-valued support function. Unobservable.

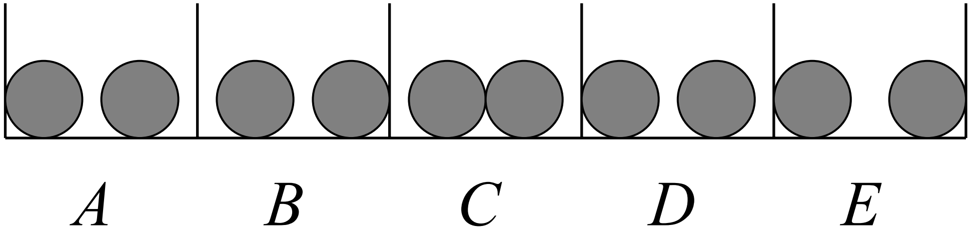
Another Strategy: The Principle of Insufficient Reason

(Leibniz, 1678; Laplace, 1776)

"If we see no reason why one case should happen more than the other then probability should be defined in terms of a ratio among cases"

E.g., horse race with five unfamiliar horses

$$P(A) = P(B) = P(C) = P(D) = P(E) = 1/5$$



Anchoring on the Ignorance Prior

What is the probability of rain tomorrow?

1. Partition state space

Rain

No Rain

2. Start automatically/unconsciously with equal probabilities

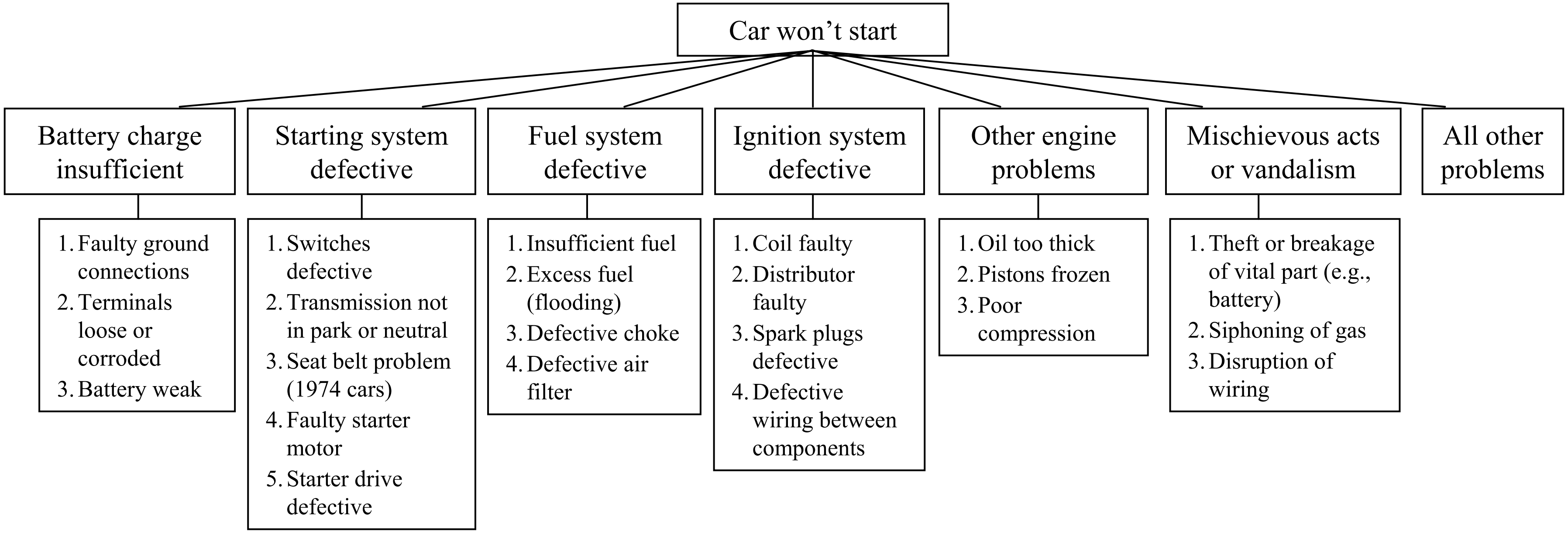
P(Rain | no thought) = 0.50

3. Adjust according to knowledge, evidence, judged support

P(Rain | thought)
= λ (support) + (1- λ) 0.50

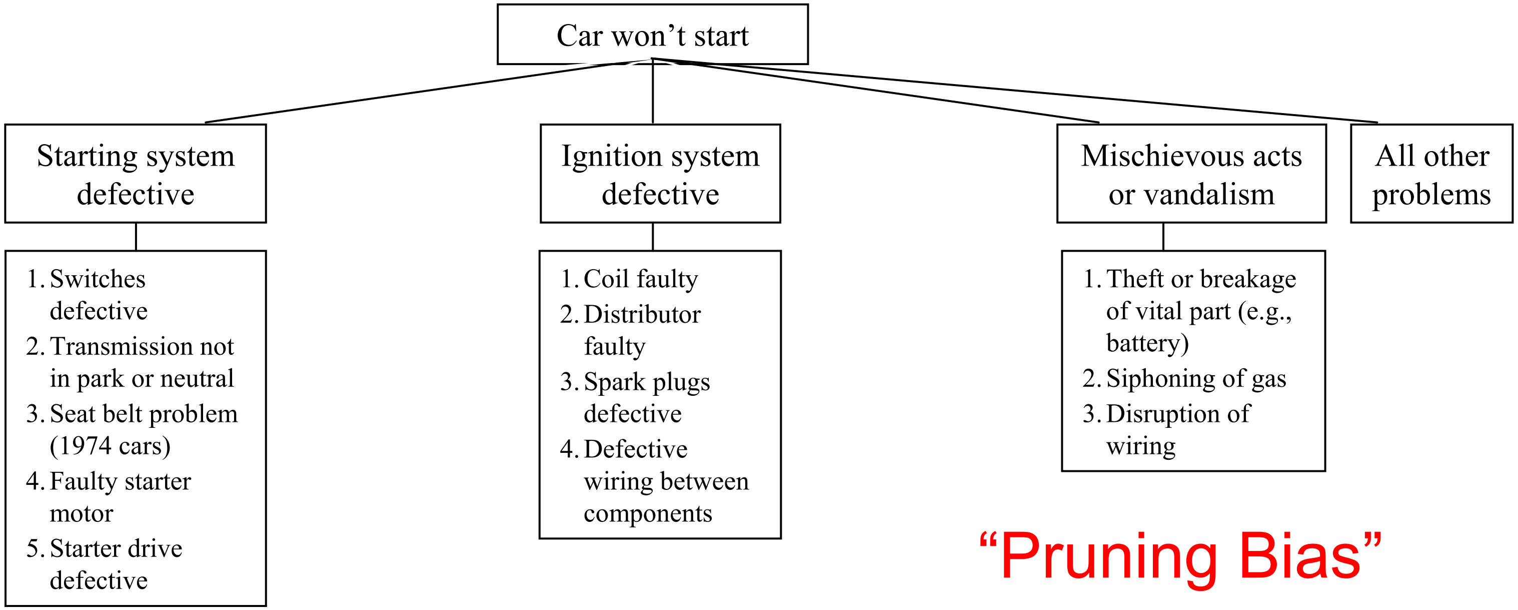
Adjustment is usually insufficient → *Partition dependence*

Fischhoff, Slovic, Lichtenstein (1978). Fault trees. *J Exp Psych.*



Car won’t start

- Battery charge insufficient
  1. Faulty ground connections
  2. Terminals loose or corroded
  3. Battery weak
- Starting system defective
  1. Switches defective
  2. Transmission not in park or neutral
  3. Seat belt problem (1974 cars)
  4. Faulty starter motor
  5. Starter drive defective
- Fuel system defective
  1. Insufficient fuel
  2. Excess fuel (flooding)
  3. Defective choke
  4. Defective air filter
- Ignition system defective
  1. Coil faulty
  2. Distributor faulty
  3. Spark plugs defective
  4. Defective wiring between components
- Other engine problems
  1. Oil too thick
  2. Pistons frozen
  3. Poor compression
- Mischievous acts or vandalism
  1. Theft or breakage of vital part (e.g., battery)
  2. Siphoning of gas
  3. Disruption of wiring
- All other problems



Car won’t start

- Starting system defective
  1. Switches defective
  2. Transmission not in park or neutral
  3. Seat belt problem (1974 cars)
  4. Faulty starter motor
  5. Starter drive defective
- Ignition system defective
  1. Coil faulty
  2. Distributor faulty
  3. Spark plugs defective
  4. Defective wiring between components
- Mischievous acts or vandalism
  1. Theft or breakage of vital part (e.g., battery)
  2. Siphoning of gas
  3. Disruption of wiring
- All other problems

“Pruning Bias”

Explanations for Pruning Bias

Availability: "Out of sight is out of mind"

Fischhoff, Slovic & Lichtenstein (1978), Russo & Kolzow (1994)
Ofir (2000), Van der Plight et al (1987)

Ambiguity: "What category does this event belong to?"

Hirt & Castellan (1988)

Credibility: "Each specified event must have some non-trivial probability."

(An example of a *demand effect*: "What is the analyst looking for?") Dube-Rioux & Russo (1988)

Our five studies

- Duke MBA students
- Some results:
  - Not just availability!
  - "Ignorance" → strong partition dependence
  - Knowledge reduces the effect
  - Not just a "demand effect"
  - Even experts are susceptible

Study 1: Categorical Partitions (Fault Tree)

A

A school other than Wharton

Wharton

0.40 “Packed”

B

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

Wharton

0.40 “Unpacked”

C

Chicago

Harvard

Kellogg

Stanford

Wharton

None of the above

0.70* “Separate Evaluation”

*$p = 0.005$ against Packed
$p = 0.05$ against Unpacked

Availability cannot be the whole story.

Study 2: Judgment Under Ignorance

246 Duke MBA students judged future close of JSX

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

	median P	% take \$30
$P(A \text{ or } B) =$	0.67	55
$P(a \text{ or } b) =$	0.30 p=0.02	31 p=0.006
$P(\mathrm{C}) =$	0.25	28
$P(c \text{ or } d \text{ or } e \text{ or } f) =$	0.60 p=0.001	58 p=0.001

Study 3: Manipulating Knowledge

Duke MBA starting salary

- $55,000 or less
- $55,001-$65,000
- $65,001-$75,000
- $75,001-$85,000
- More than $85,000

- $85,000 or less
- $85,001-$95,000
- $95,001-$105,000
- $105,001-$115,000
- More than $115,000

Harvard Law starting salary

- $60,000 or less
- $60,001-$70,000
- $70,001-$80,000
- $80,001-$90,000
- More than $90,000

- $90,000 or less
- $90,001-$105,000
- $105,001-$115,000
- $115,001-$130,000
- More than $130,000

Study 3 Results

Duke MBA students in decision analysis elective (n=120)

Median knowledge of Duke MBA salaries: 7/10

Median knowledge of Harvard Law salaries: 2/10

	Duke MBA		Harvard Law	
	< $85K	≥ $85K	< $90 K	≥ $90K
Low	**0.75**	0.25	**0.75**	0.25
High	0.40	**0.60**	0.30	**0.70**

= packed
bold = separate evaluation

Overall test for partition dependence: $p<0.0001$

Less knowledge ➔ Stronger effect

But barely significant: $p=0.05$ *(t-test)*

Study 4: Does the Partition Convey Information?

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:_______________

Below 1000 _______
1000-2000 _______
1001-4000 _______
Above 4000 _______

Index: _________________

4000 or below _______
4001-8000 _______
8001-16000 _______
Above 16000 _______

FTSE 100 Stock Index
Closing value, Oct 21, 2005

Group 1

Assess probabilities
for these events:

- FTSE ≤ 4200
- 4200 < FTSE ≤ 4800
- 4800 < FTSE ≤ 5500
- 5500 < FTSE ≤ 6300
- 6300 < FTSE

Group 2

Assess probabilities
for these events:

- FTSE ≤ 2800
- 2800 < FTSE ≤ 3200
- 3200 < FTSE ≤ 3700
- 3700 < FTSE ≤ 4200
- 4200 ≤ FTSE

Study 4: Results: JSX

INDEX: *JSX*

Below 1000	_ 0.25_
1000-2000	_ 0.25_
2001-4000	_ 0.25_
Above 4000	__0.25_

INDEX : *JSX*

Below 4000	__0.25_
4000-8000	__0.25_
8000-16000	__0.25_
Above 16000	__0.25_

Median knowledge rating: *0/10*

Study 4 Results: NASDAQ

INDEX: *NASDAQ*

Below 1000 __0.00_

1000-2000 __0.05_

2001-4000 __0.30_

Above 4000 __0.50_

INDEX: *NASDAQ*

Below 4000 __0.25_

4000-8000 __0.50_

8000-16000 __0.05_

Above 16000 __0.00_

Median knowledge rating: *7/10*

Study 4 Results Summary

Weekend Executive MBA students (n = 102)

	NASDAQ		***JSX***	
	≤4000	*>4000*	*≤ 4000*	*>4000*
Low	**0.50**	0.50	**0.75**	0.25
High	0.25	**0.75**	0.25	**0.75**

= packed
bold = *separate evaluation*

Overall test for partition dependence: $p<0.0001$

NOTE: Partition dependence disappears for "experts" on NASDAQ (knowledge ≥ 7)

Can't attribute effect to information content of partitions (Grice 1975)

Study 5: Probability Assessment Experts

Survey: Members on the Decision Analysis Society email list ($n = 55$)

- Approached by e-mail
- 86% have Ph.D.
- 75% have taught course in Decision Analysis
- 63% have elicited probabilities for applied DA project in last 2 years
- Total of 156 projects over previous two years

Study 5: Method

What is the last digit of your local telephone number?
If this number is *even*, please write “DAS” in the space provided above the tree on the *left* and “SQA” in the space provided above the tree on the *right*.

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

Society:_______________

≤400 _______
401-600 _______
601-800 _______
801-1000 _______
>1000 _______

Society: _________________

≤1000 _______
1001-1200 _______
1201-1400 _______
1401-1600 _______
>1600 _______

Study 5 Results

DAS email list (n=55)

	DAS		*SQA*	
	≤1000	*>1000*	*≤ 1000*	*>1000*
Low	**0.90**	0.10	**0.80**	0.20
High	0.65	**0.35**	0.55	**0.45**

= *packed*

bold = *separate evaluation*

Overall test for partition dependence: $p<0.0001$

NOTE: No knowledge effect

Even we are susceptible!

Study 5: Super Experts

25 Experienced Decision Analysts

- Ph.D. in 1985 or earlier AND
- At least one applied project in last 2 years AND
- Taught at least one course in Decision Analysis

	DAS		*SQA*	
	≤1000	*>1000*	*≤ 1000*	*>1000*
Low	**0.88**	0.12	**0.80**	0.20
High	0.75	**0.25**	0.58	**0.42**

= *packed*

bold = *separate evaluation*

Overall test for partition dependence: $p=0.05$

"You know who you are"

How big is the problem?

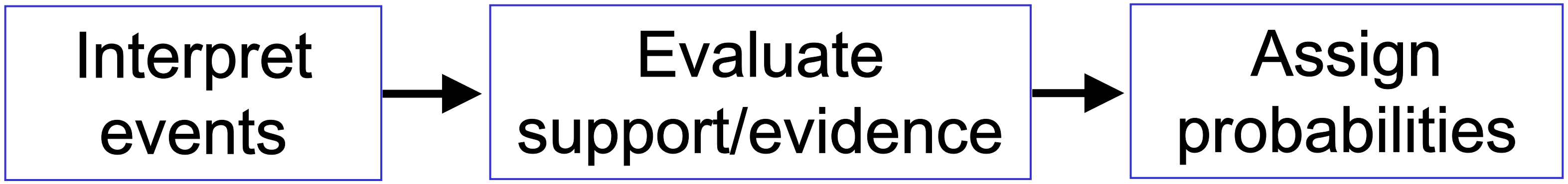
What methods do we use to assess to assess continuous probability distributions?

Methods used:

- **Ask for probabilities of pre-specified intervals: 58% of the time**
- **Ask for percentiles of a distribution: 56% of the time**

Some Implications for Practice

Stages in Probability Assessment



Interpret events → Evaluate support/evidence → Assign probabilities

Mechanisms	*Availability, Ambiguity*	*Judgmental heuristics, credibility assumptions*	*Ignorance prior*
How to counteract	• Clarity test • Careful conditioning of experts	• Develop partition with expert. • Expert articulates reasoning, assumption, information sources	• Careful specification of partition • Explore state space "evenhandedly" • Use multiple partitions
	Existing Best Practice		New

What about using a model to debias judgments?

Imagine a state space partitioned into k events

Multiplicative model from Fox & Rottenstreich (2003):

$$R(A,\overline{A}) = \left[\frac{1}{k-1}\right]^{\lambda}\left[\frac{s(A)}{s(\overline{A})}\right]^{1-\lambda}$$

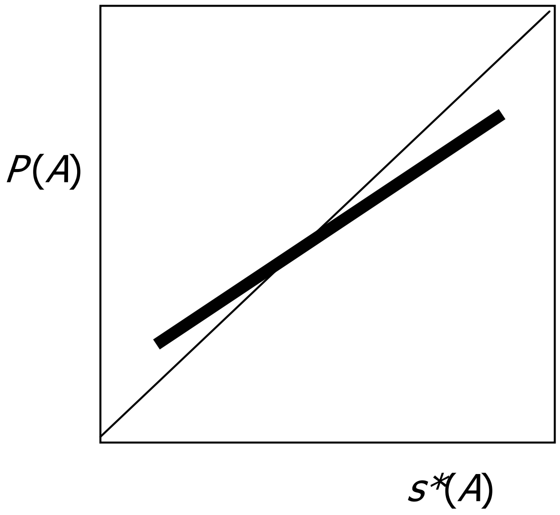
Linear model from Clemen & Ulu (2005):

$$P(A) = \alpha\, s^*(A) + (1\text{-}\alpha)(1/k)$$

α and λ are weighting parameters.

Linear Model

$$\boldsymbol{P(A) = \alpha\, s^*(A) + (1\text{-}\alpha)(1/k)}$$



$$0 \leq \alpha \leq 1$$

$$s^* \geq 0,\ \Sigma s^* = 1,\ s^*(A \cup B) = s^*(A) + s^*(B)$$

Interpretations of *s**

- Normalized support
- “True” or “covert” probability

A Bayesian Interpretation

Bayesian updating of ignorance prior based on recruited support *s**.

Model: Judge has a prior for an unknown probability (parameter of a Bernoulli process), treats recruited support as equivalent sample information.

$$P(A) = \alpha\, s^*(A) + \underbrace{(1-\ \alpha)}_{\beta}\ (1/k)$$

α is weight given to equivalent sample information (based on amount of information)

Support s^* viewed as equivalent sample information

Ignorance prior based on A being one of k elements of the partition.

Some properties of the model

Known properties of subjective probabilities:

- **Binary complementarity**

  Complementary probabilities sum to one

- **Subadditivity of directly-assessed probabilities increases with size of partition.**

  Subjective probabilities sum to more than one

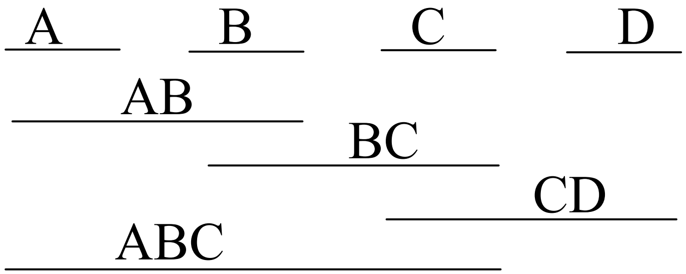
- **Ignorance-prior effect**

  The less knowledge, the closer the probabilities are to equality

New properties

- **Interior additivity**

  $P'_2(A) = P_{AB} - P_B = P_{AC} - P_C$ for any B, C
  (no null or sure events)



$$P(\mathrm{A}) = P_{\mathrm{A}} = P_{\mathrm{AB}} - P_{\mathrm{B}}$$
$$= P_{\mathrm{ABC}} - P_{\mathrm{BC}}$$

IA:

$P(A) \equiv P(A \cup B) - P(B)$

- For all disjoint events or "hypotheses" A, B
- "Interior" only:
  - $A, B \neq \varnothing$
  - $A \cup B \neq$ sure event

Empirical observation of IA by Wu and Gonzalez (1999)

New properties

- **Interior additivity**

  $P'_2(A) = P_{AB} - P_B = P_{AC} - P_C$ for any B, C
  (no null or sure events)

- **Indirect probabilities are less than direct probabilities**

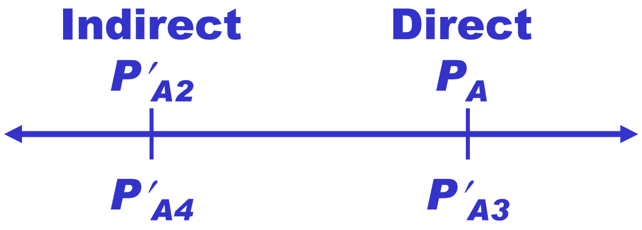
  $$\begin{aligned} P'_2(A) = P_{AB} - P_B = &\ \alpha\,[s^*(AB) - s^*(B)] = \alpha s^*(A) \\ \leq &\ \alpha s^*(A) + \beta \ = P_A \end{aligned}$$

- **Superadditivity of indirect probabilities**

  Indirect probabilities sum to less than 1.

- **"Flip-flop" effect**

  $P'_3(A) = P_{AB} + P_{BC} - P_{ABC} = P_A$
  $P'_4(A) = P_{AB} + P_{CD} - P_{BC} - P_D = P'_2(A)$



Indirect **Direct**

P'_{A2} **P_A**

P'_{A4} **P'_{A3}**

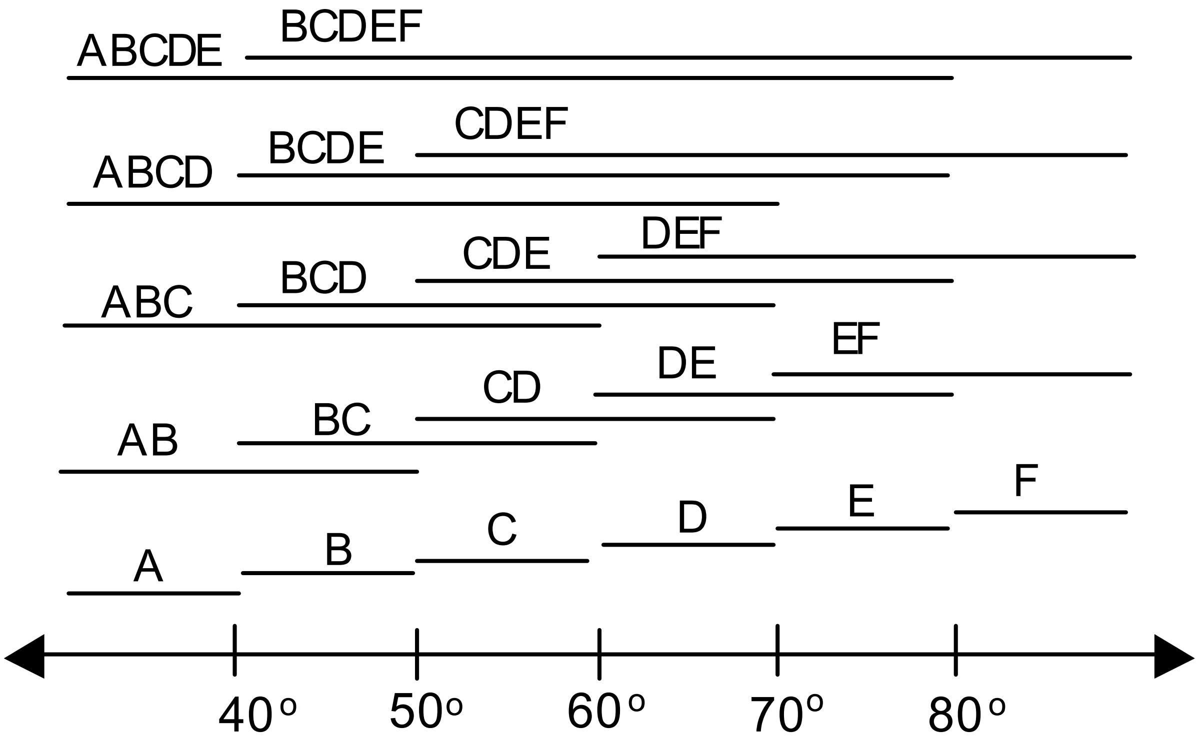
Summary

H1a	Interior additivity of P'_2	
H1b	$\sum P'_i = K$. Sum of indirect probs constant for all assessment structures.	
H2	Indirect probabilities are less than direct probabilities	
H3	Indirect probabilities add to less than one (superadditivity)	
H4a	Flip-Flop effect	For 3-element probs
H4b		For 4-element probs

Data from Tversky & Fox (1995)

- Probability assessments for
  - 1991 NBA games: Trailblazers vs Jazz
  - 1992 Superbowl: Bills vs Redskins
  - Dow Jones Industrial Average the following week
  - High temperatures in San Francisco & Beijing

Many assessments for each variable



20 different assessments for temperature. Similar for others.

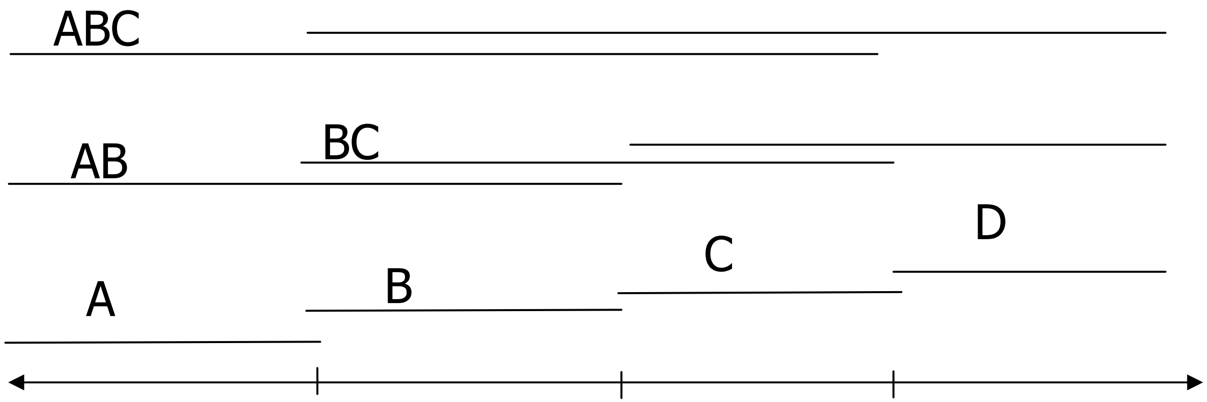
Aggregate results - TF

Hypothesis	Test	# of tests	# of rejections for significance level 0.001	0.01	0.05	0.20
H1a: Interior Additivity	Wilcoxon	292	2	16	25	33
H1b: IA – Constant sum	Wilcoxon	50	3	5	10	14
H2: Indirect < Direct	Wilcoxon	262	163	215	241	249
H3: Superadditivity	Sign	107	91	100	104	107
H4a: Flip-flop for 3	Wilcoxon	64	5	6	8	14
H4b: Flip-flop for 4	Wilcoxon	52	2	7	9	17

Two new studies

- TF assessment conditions:
  - No constraints imposed
  - Probabilities assessed one at a time, not the entire partition.
- In our studies
  - Constrain $P_{AB} > P_A$ to ensure positive indirect probabilities
  - Assessed all probabilities at the same time
  - More similar to realistic decision or risk analysis
  - Expect smaller effect for
    - H2 (indirect<direct)
    - H3 (superadditivity of indirect)

Assessment tasks

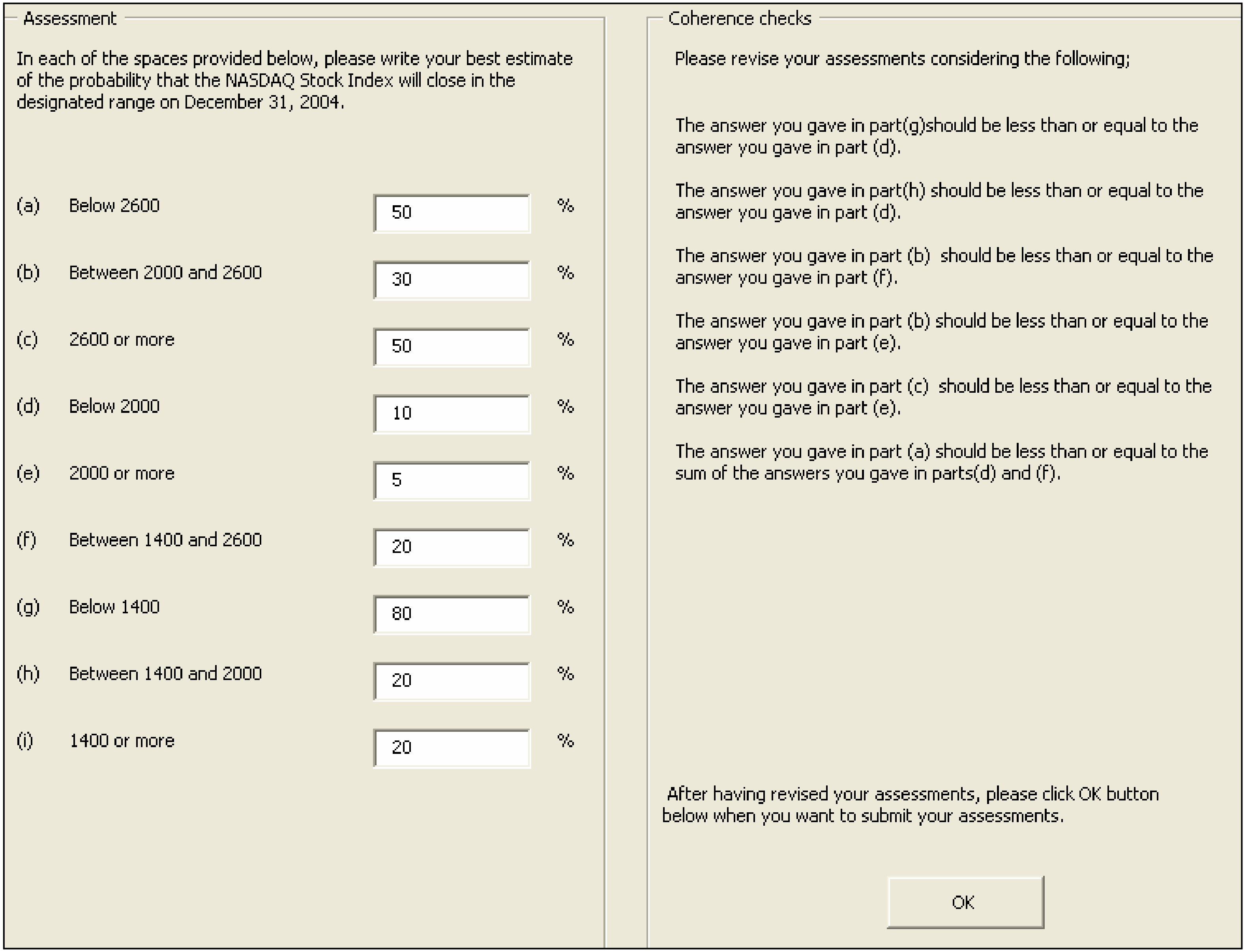


Durham temp.	60	70	80
Nasdaq Index	1400	2000	2600
MBA salary	45,000	85,000	125,000
GPA	1.5	2.5	3.5
Rent	600	800	1000

Questions asked

- "In each of the spaces provided below, please write your best estimate of the probability that the high temperature in Durham on March 1, 2003, will fall in the designated range.“
- "In each of the spaces provided below, please write your best estimate of the probability that the NASDAQ Stock Index will close in the designated range on December 31, 2004."
- "In each of the spaces provided below, please write your best estimate of the probability that the salary of a 2003 Fuqua School of Business MBA graduate will fall in the designated range."
- "In each of the spaces provided below, please write your best estimate of the probability that the grade point average (GPA) of a random student from 2004 Fuqua MBA class will fall in the designated range at the end of Spring 2003."
- "In each of the spaces provided below, please write your best estimate of the probability that the monthly rent of an apartment with 2 bedrooms within 2 miles of the Fuqua School of Business will fall in the designated range."

Example assessment screen. $n_1 = 77$, $n_2 = 118$



Assessment

In each of the spaces provided below, please write your best estimate of the probability that the NASDAQ Stock Index will close in the designated range on December 31, 2004.

(a) Below 2600 50 %
(b) Between 2000 and 2600 30 %
(c) 2600 or more 50 %
(d) Below 2000 10 %
(e) 2000 or more 5 %
(f) Between 1400 and 2600 20 %
(g) Below 1400 80 %
(h) Between 1400 and 2000 20 %
(i) 1400 or more 20 %

Coherence checks

Please revise your assessments considering the following;

The answer you gave in part(g)should be less than or equal to the answer you gave in part (d).

The answer you gave in part(h) should be less than or equal to the answer you gave in part (d).

The answer you gave in part (b) should be less than or equal to the answer you gave in part (f).

The answer you gave in part (b) should be less than or equal to the answer you gave in part (e).

The answer you gave in part (c) should be less than or equal to the answer you gave in part (e).

The answer you gave in part (a) should be less than or equal to the sum of the answers you gave in parts(d) and (f).

After having revised your assessments, please click OK button below when you want to submit your assessments.

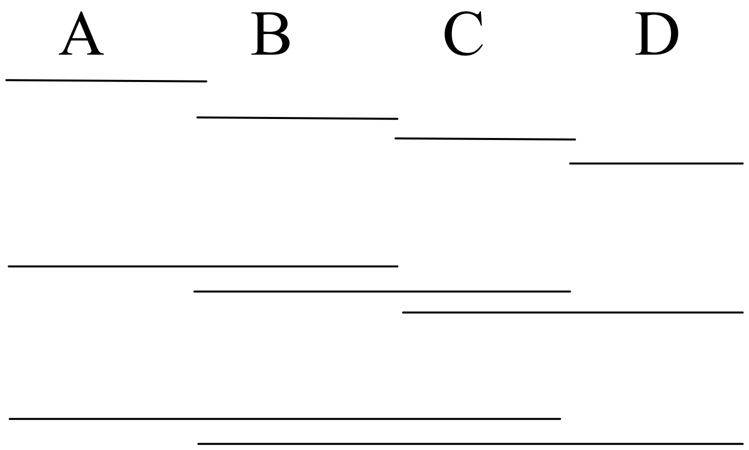
Aggregate results

			# of rejections for significance level			
Hypothesis	**Test**	**# of tests**	**0.001**	**0.01**	**0.05**	**0.20**
H1a: Interior Additivity	Wilcoxon	90	0	0	3	9
H1b: IA – Constant sum	Wilcoxon	10	0	0	2	3
H2: Indirect < Direct	Wilcoxon	140	20	38	48	64
H3: Superadditivity	Sign	30	10	14	17	21
H4a: Flip-flop for 3	Wilcoxon	20	0	0	2	3

How to recover $s*$

1. Partition the state space.
2. Assess probabilities.
3. Calculate indirect probabilities $P'_2(A) = P_{AB} - P_B$ as needed.
4. Normalize to get

$$\begin{aligned} P^*(A) &= P'_2(A)/\Sigma P'_2 \\ &= \alpha s^*(A)/\alpha\, \Sigma s^* \\ &= s^*(A) \end{aligned}$$



P* = s* represents the judge's knowledge, "untainted" by the ignorance prior.

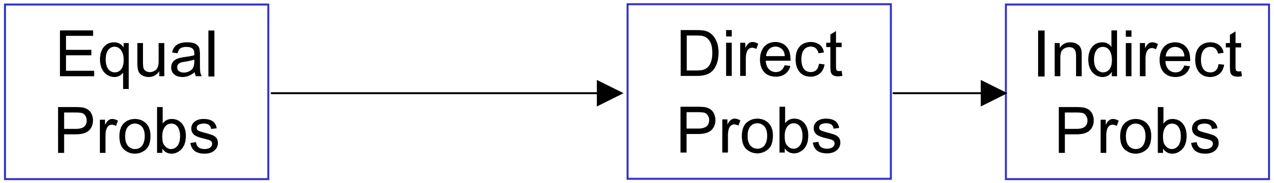
Does it work?

- Study 3 to test performance of the procedure
- n_3 = 85 grad students
- Each one made interval assessments for four different variables
  - Two conditions: Direct and Indirect
  - 4 questions taken from a pool of 100 (half familiar "everyday" questions, half almanac)
  - Assessment intervals created randomly
  - Knowledge ratings on the variables

Results

Difference from “ignorance prior”

Indirect probabilities are farther away (statistically significant)



Equal Probs → Direct Probs → Indirect Probs

Accuracy (Brier Score)

Inconclusive results

Indirect Probs perform slightly better for high-knowledge questions.

Results

Probability scores, decomposed into reliability (calibration) and resolution:

		Indirect	Direct
All data	Base score	0.730	0.727
	Reliability	0.530	0.500
	Resolution	0.447	0.414
	Total	**0.818**	**0.813**
Low knowledge	Base score	0.743	0.721
	Reliability	0.649	0.584
	Resolution	0.571	0.525
	Total	**0.821**	**0.780**
High knowledge	Base score	0.701	0.733
	Reliability	0.629	0.665
	Resolution	0.515	0.549
	Total	**0.815**	**0.848**

Summary

- Partition dependence is a real and insidious phenomenon.
- Linear probability model is consistent with support theory, partition dependence, interior additivity.
    - Evidence for the model is strong.
- Sharp prescriptive implication: Normalize indirect probabilities.
    - Performance test shows
        - Indirect probs farther from ignorance prior
        - Little difference in average score
- Decision analysis beginning to use behavioral results to develop sharp prescriptive procedures.
    - Stay tuned for new developments!