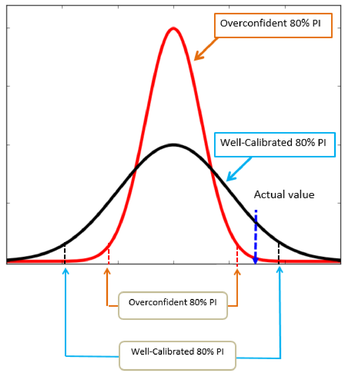
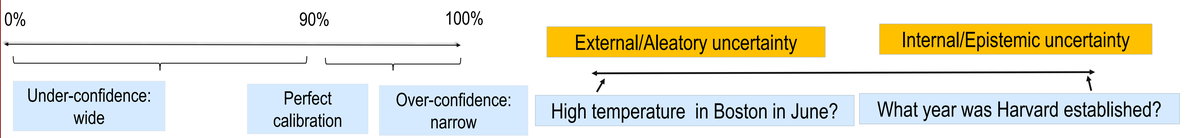
Differential Miscalibration of Subjective Prediction Intervals for Events Involving External (Aleatory) and Internal (Epistemic) Uncertainties

David Budescu Department of Psychology, Fordham University, New York
Saemi Park Department of Psychology, The Ohio State University, Ohio

1. Introduction

- Judges often fail to adjust their interval estimates appropriately to match the prescribed confidence levels.
- Their subjective intervals (PI) tend to be too **narrow** suggesting **Overconfidence**.
- Budescu and Du (2007) show that judges are sensitive to the prescribed confidence levels in **within-subject** designs by generating multiple CIs.
- Teigen & Jorgenson (2005) speculate that judges would be better calibrated when generating intervals about events involving **external uncertainty**.
- We tests whether (a) the insensitivity to confidence levels varies across the two types of uncertainty and (b) elicitation procedures that requires multiple judgments are superior to one–shot elicitations.



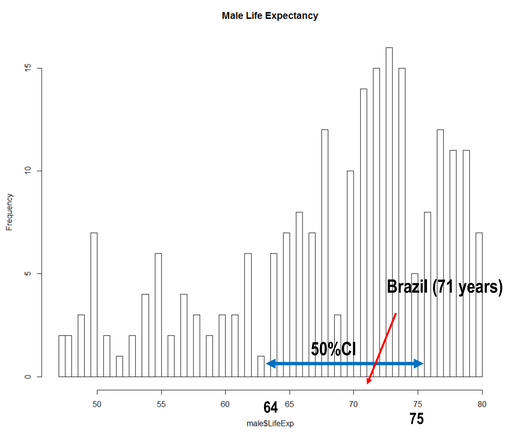


2. Method

1. Participants: 192 recruited from Amazon Mturk
2. Questions: We elicited 3 points: Lower & Upper bounds and Best estimate for 20 questions
   - **External/Aleatory**: Distributions from 3 well-defined domains (193 UN countries, 51 US states, 52 weeks in a year)

   EXAMPLE) 50%PI of Male life expectancy across 193 UN countries in 2012
   - **Internal/Epistemic**: An event is selected in each case to be as close as to possible to the Median.

   EXAMPLE) 50%PI of Male life expectancy in Brazil
3. Details: (a) Min and Max provided (b)Elicited 50% and 90% PIs (some control)
4. Measures of Performance



External/Aleatory Uncertainty	Internal/Epistemic Uncertainty
1. Median ratio: (UB-LB)/Actual width across all items	**1. Hit Rate(HR)** : (# of items bracketing the actual value/n)*100
2. Width of interval: (UB-LB)/Actual median	**2. Relative width of interval**: (UB-LB)/Actual value
3. Coverage rate: %of data points bracketed by the PI	**3. Q-Score**: measures that combines HR and the width of interval, ideally 0.
4. Absolute relative bias: ARB= |(Best- Median)/Median|	**4. Absolute relative bias**: ARB= |(Best-Actual)/Actual value|

3. Results

1. Can judges differentiate between various levels of confidence when providing a single PI? (Between-subjects)

External/Aleatory Uncertainty

Dependent variable	Mean (SD) of 90%	Mean (SD) of 50%
Median Ratio	**1.01 (.19)**	**1.97 (.45)**
Median Width	.57 (.17)	.57 (.20)
Coverage rate	76.61 (14.36)	73.42 (15.82)
Absolute relative bias	.27 (.12)	.28 (.12)

- **Median ratio***: well-calibrated 90%PIs and underconfident 50%PIs
- All other measures nearly identical

Internal/Epistemic Uncertainty

Dependent variable	Mean (SD) of 90%	Mean (SD) of 50%
Hit rate	86.39 (13.13)	82.74 (15.73)
Median Width	.40 (.14)	.38 (.18)
Q-score	**-.14 (.09)**	**-.41 (.20)**
Absolute relative bias	.36 (.17)	.30 (.12)

- HR: 90%PI slightly overconfident and 50%PI widely
- **Q-score***: significant, 90%PIs are superior
- All other measures nearly identical

2. Do judges learn to adjust PIs to differentiate properly levels of confidence when providing multiple PIs?

External/Aleatory Uncertainty

Dep. Variable	Condition1 (90%/50%)	Condition2 (50%/90%)	Condition3 (90%/90%)	Condition4 (50%/50%)	Tukey's HSD
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	
Median Ratio	.82 (.49)	-.93 (.40)	-.03 (.08)	.01 (.17)	2<3=4<1
Width of interval	-.09 (.16)	.05 (.19)	-.02 (.11)	-.01 (.07)	1<2, 1<3=4=2
Coverage rate	-11.88 (14.75)	6.78 (15.15)	-1.36 (4.84)	.79 (7.33)	1<2, 1<3=4=2
ARB	-.01 (.15)	.01 (.11)	.03 (.08)	.03 (.14)	-

- No differences in the control conditions
- **Ratio***: 50%PIs > 90%PIs
- **Width and Coverage rate***: 50%PIs < 90%PIs Adjusted in the appropriate direction

Internal/Epistemic Uncertainty

Dep. Variable	Condition5 (90-50%)	Condition6 (50-90%)	Condition7 (90-90%)	Condition8 (50-50%)	Tukey's HSD
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	
Hit rate	-7.19 (12.87)	6.85(10.86)	1.02 (11.59)	-2.13 (11.58)	5<6, 5=8=7, 8=7=6
Width of interval	-.34 (.70)	.37 (.60)	-.07 (.20)	.02 (.15)	5<7=8<6
Q-score	-.25 (.18)	.30 (.21)	.01 (.13)	-.01 (.07)	5<8=7<6
ARB	-.05 (.12)	.01 (.12)	.00 (.14)	.00 (.08)	-

- **HR***: 90% > 50%
- **Width of interval and Q-score*** : 90% > 50%

3-1. Are there differences between PIs based on external and internal uncertainties?

Uncertainty	Normalized absolute Bias			Normalized interval width		
	Prescribed Level of Confidence					
	50%	90%	Mean	50%	90%	Mean
External	.14 (.03)	.14 (.04)	.14 (.04)	.76 (.17)	.80 (.15)	.78 (.16)
Internal	.14 (.04)	.15 (.04)	.15 (.04)	.62 (.23)	.70 (.22)	.66 (.22)
Mean	.14 (.04)	.14 (.04)	.14 (.04)	.69 (.21)	.75 (.19)	.72 (.20)

- No significant difference in Bias.
- **Width of 90%PIs >50%PIs.**
- **Width of Internal PIs < External PIs**

3-2. Are there differences between PIs based on external and internal uncertainties?

Uncertainty	Hit rate			Difference between hit rate and prescribed confidence		
	Prescribed Level of Confidence			Prescribed Level of Confidence		
	50%	90%	Total	50%	90%	Total
External	76.10 (08.51)	76.75 (10.50)	76.42 (09.52)	26.10 (08.52)	-13.25 (10.50)	06.43 (21.94)
Intrenal	82.74 (15.73)	86.39 (13.13)	84.57 (14.53)	32.74 (15.73)	-3.61 (13.13)	14.57 (23.27)
Mean	79.42 (13.02)	81.57 (12.78)	80.50 (12.91)	29.42 (13.02)	-8.43 (12.78)	10.50 (22.92)

- Pseudo HR: % of cases bracketing the median value used in the epistemic case
- HR of 90%PIs are (non-significantly) slightly higher
- **HR of PIs based on Internal uncertainty is significantly higher**

4. Discussion

- The degree of sensitivity to the prescribed confidence levels will be more pronounced in within-subject settings **SUPPORTED!**
- Judges will be more sensitive to the prescribed level of confidence when providing PIs based on external uncertainties. **NOT SUPPORTED**
- The PIs based on External/aleatory uncertainty are wider, reflecting higher degree of uncertainty