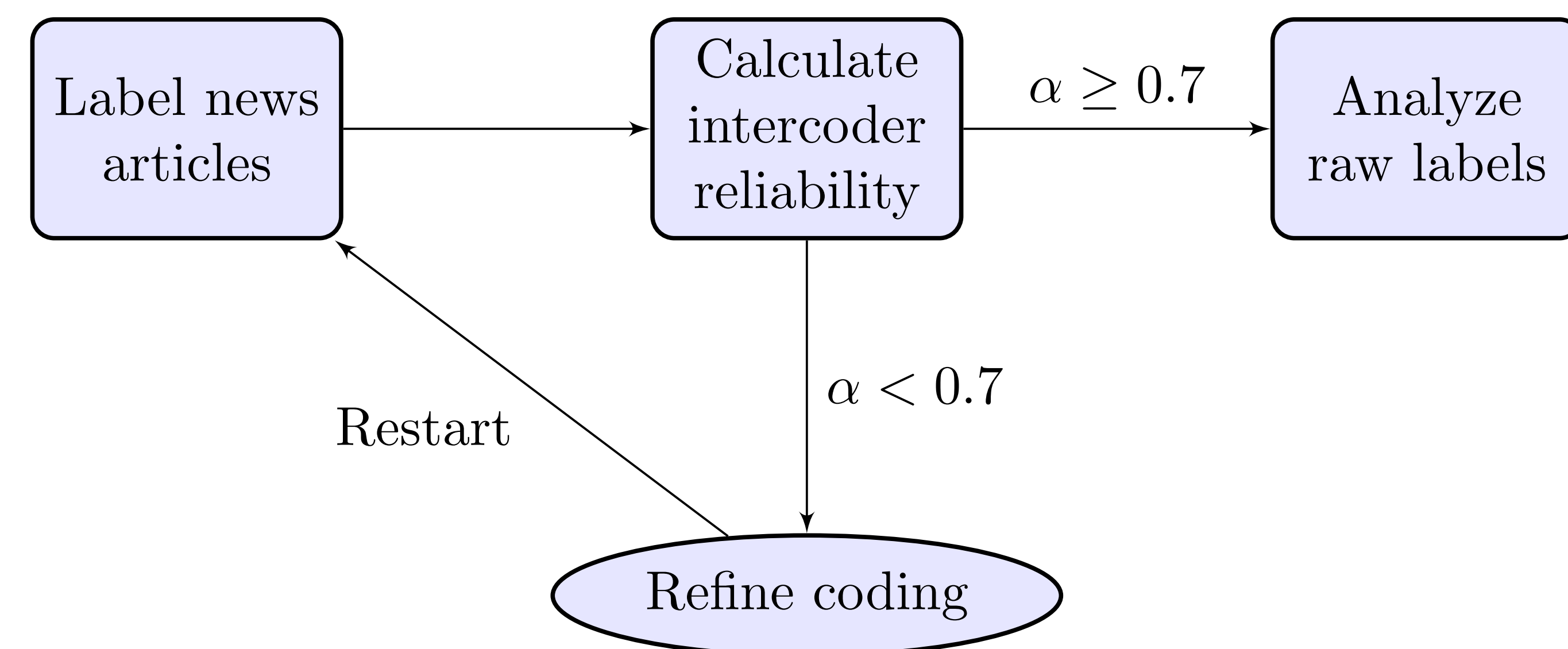


Getting the Most Out of Human Coders with Statistical Models

Matthew Tyler (Stanford University)

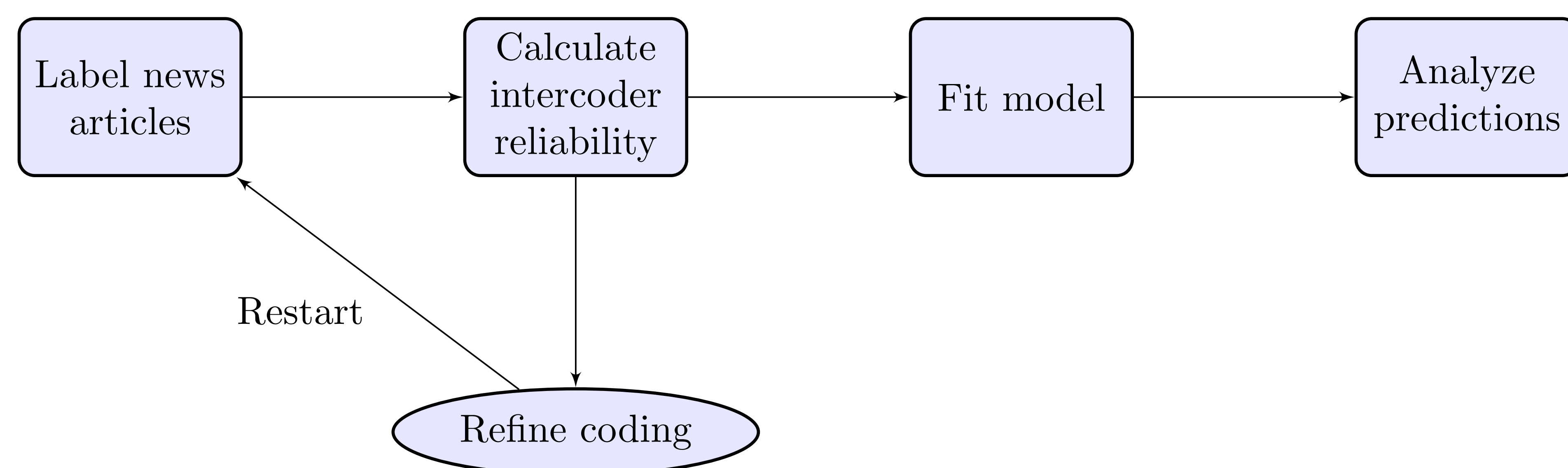
Paper link: <https://forms.gle/D2iMzBATEoHbnppT9>

STATUS QUO Standard Coding Procedure (SCP)



- Refine coding until arbitrary threshold met
- **Coders trusted implicitly, but raw labels can still have error**
- **Measurement error spoils final analysis**
- All coders treated equally

PROPOSAL Model-Based Coding Procedure (MCP)



- Refine coding until no longer cost-justified
- Model yields: $P[\text{True Label} \mid \text{Coder Labels}]$; **captures uncertainty**
- # Coders ≥ 3 : can capture heterogeneous coder accuracy
- **Model gives higher weight to more accurate coders**

INTUITION Why Do Models Work?

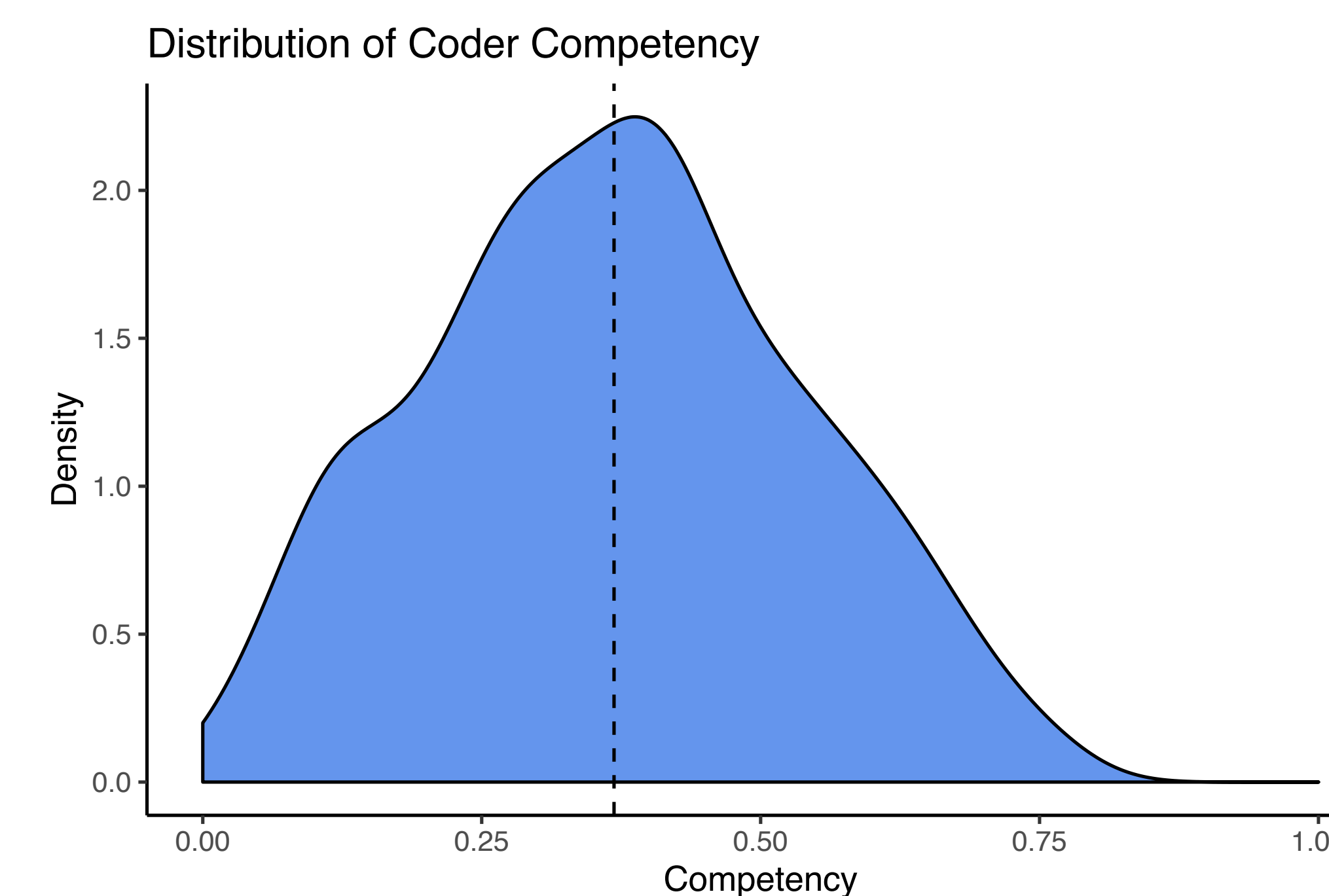
- If two coders agree, either both right or both wrong
- Let $P(\text{Coder } j \text{ Correct}) = \beta_j$
- $P(\text{Coders 1 and 2 Agree}) = P(\text{Both Correct}) + P(\text{Both Incorrect})$
- $P(\text{Coders 1 and 2 Agree}) = \beta_1\beta_2 + (1 - \beta_1)(1 - \beta_2)$
- Identify parameters with enough agreement/disagreement data

EXAMPLE MODEL Coder Competency Estimation

- Object i 's true label = $Z_i \sim \text{Cat}(\lambda)$
- Coder j 's competency = β_j
- Coder j 's guessing behavior = δ_j
- Coder j 's label for object i is X_{ij} :
 - Is the truth clear to the coder? $C_{ij} \sim \text{Bernoulli}(\beta_j)$
 - If $C_{ij} = 1, X_{ij} = Z_i$
 - If $C_{ij} = 0, X_{ij} \sim \text{Cat}(\delta_j)$

APPLICATION Coding Political News Articles

- Data: Peterson et al. (Forthcoming)
- 605 MTurkers labeled slant of 50,000 news articles (92% by one coder)
- Pro-Democratic, Neutral, or Pro-Republican
- **Example Headlines:**
- "Elected Democrat & Hillary Clinton Campaign Staffer SENT TO PRISON!"
 - 4 coders unanimously labeled Pro-Republican
 - $P(\text{Pro-Rep.}) = 0.99$
- "Donald Trump says soldiers with PTSD aren't strong"
 - 2 coders disagreed, but one coder is more accurate
 - $P(\text{Pro-Dem.}) = 0.61, P(\text{Neutral}) = 0.36$



	Probability Headline Mentions			
	Trump	Trump	Clinton	Clinton
Pro-Rep.	-0.16** (0.03)	-0.58** (0.02)	0.16** (0.03)	0.36** (0.02)
Neutral	-0.02 (0.02)	-0.42** (0.01)	-0.03** (0.01)	0.10** (0.01)
(Intercept)	0.58** (0.002)	0.88** (0.01)	0.28** (0.002)	0.15** (0.01)
Method	SCP	MCP	SCP	MCP
N	50,204	50,204	50,204	50,204

- Left: MCP uncovers **large variance** in coder accuracy ("Competency")
- Right: example final analysis: regressing topic on article slant
- Right: **MCP finds stronger link between topic and slant than SCP because it adjusts for measurement error**