Statistical Modelling in Surveys without Neglecting *The Undecided*

Julia Plass, *Paul Fink*, Nobert Schöning, Thomas Augustin

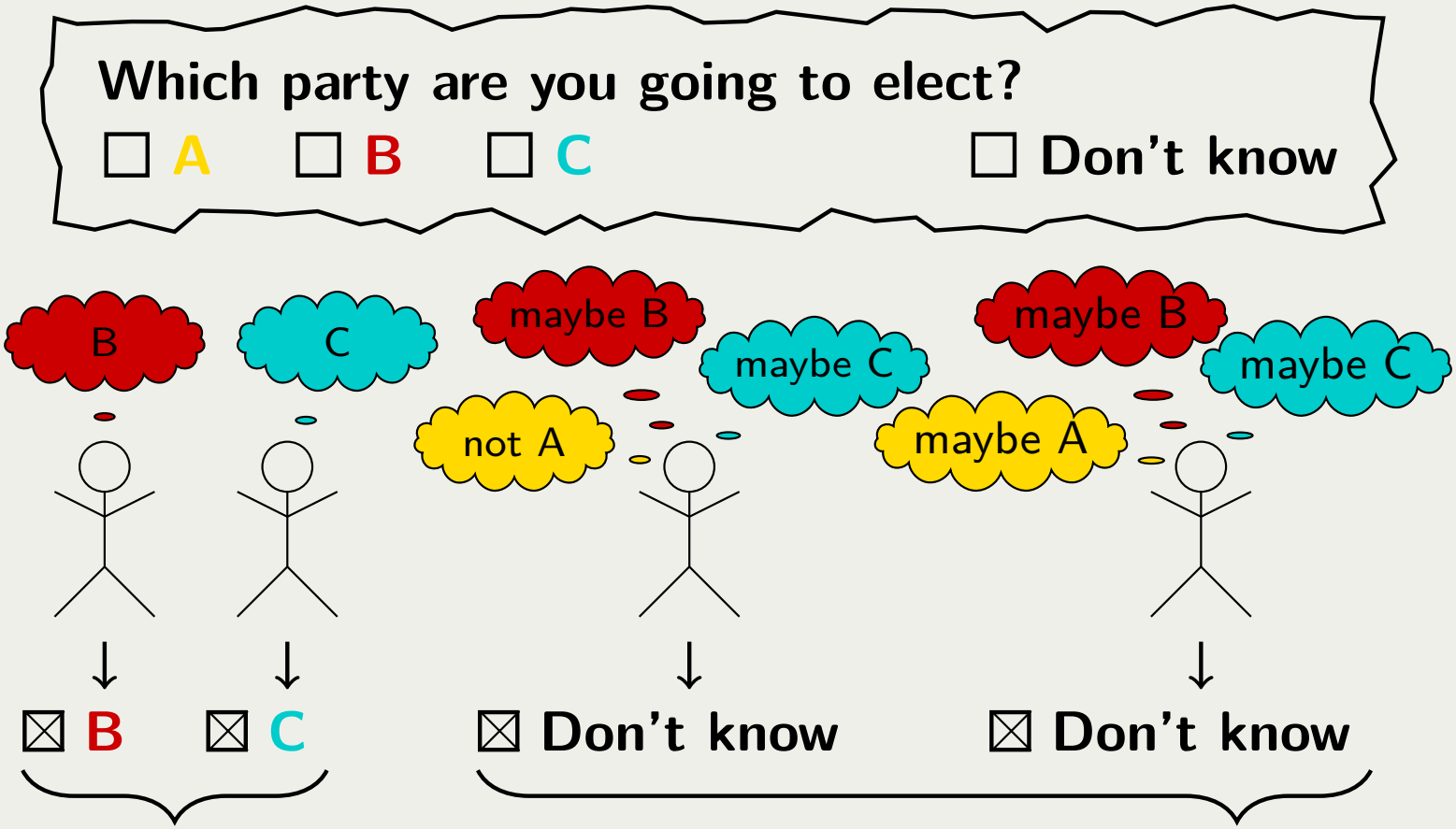
Department of Statistics & Geschwister Scholl Institute of Political Science
LMU Munich

July 22nd, 2015
ISIPTA '15 Pescara, Italy

Biography

Julia Plass	PhD student of Thomas Augustin $\rightarrow$ Presentation on Tuesday
Paul Fink	PhD student of Thomas Augustin M. Sc. in Statistics Strong interest in classification with *deficient* data
Norbert Schöning	M.A. in Political Sciences and Diploma in Sociology Dealing with missing data in studies of voting behaviour
Thomas Augustin	Head of Workgroup

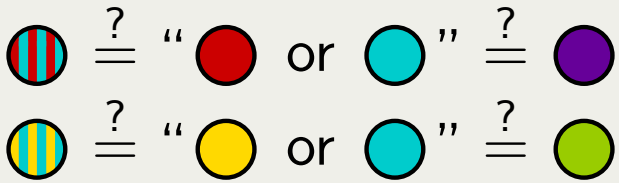
Current Analyses in Election Studies



- Only include decisive respondents
- Loss of information $\Longrightarrow$ biased results

Views on Undecided (Couso, Dubois & Sánchez, 2014)

Indecisiveness between options



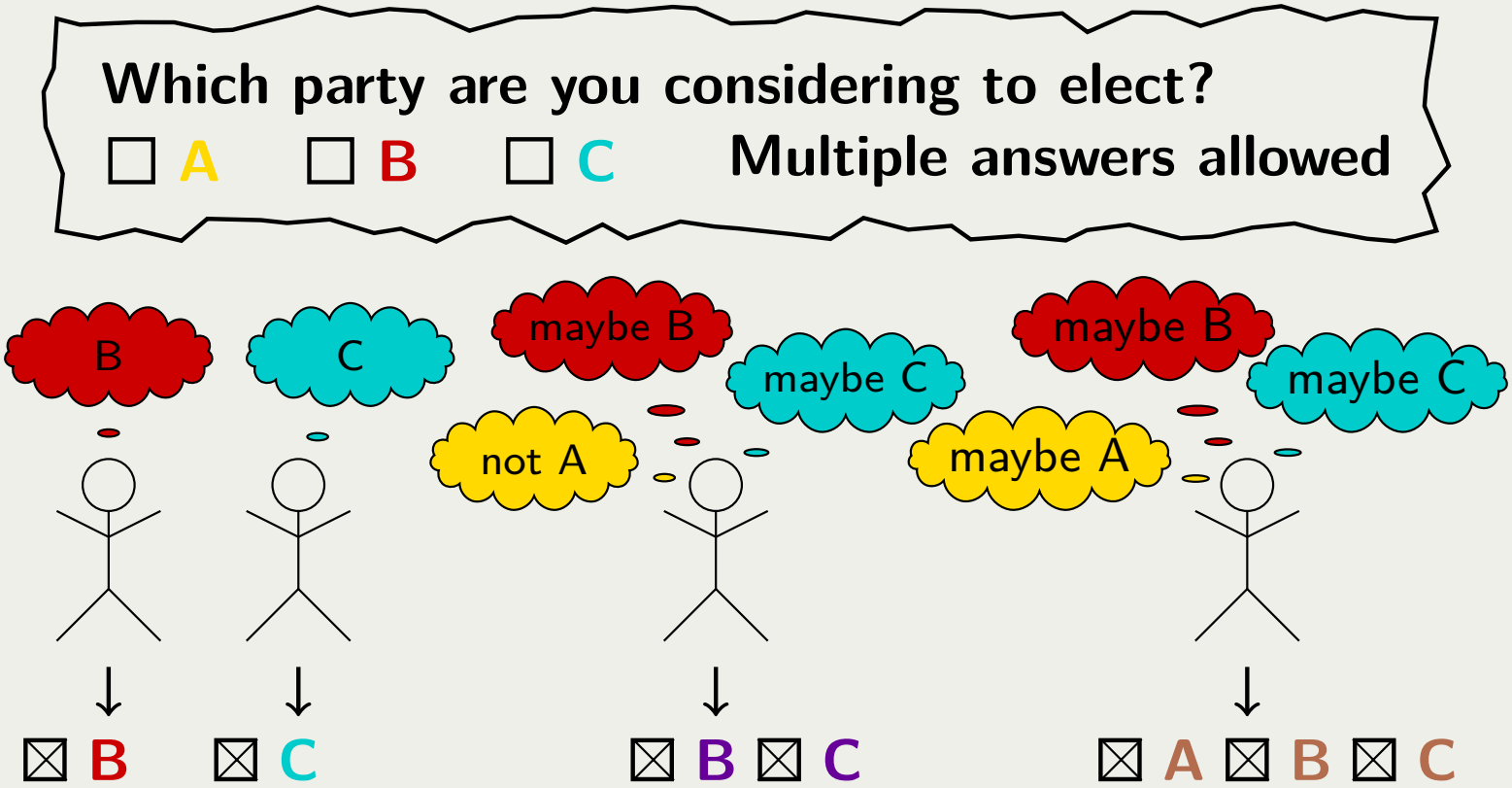
Epistemic Imprecision

- Imprecise observation of something precise
- Truth is masked by underlying coarsening mechanism

Ontic Imprecision

- Precise observation of something imprecise
- Truth is represented by coarse observations
- Respondents without unique preference in pre-election studies

Don't neglect the Undecided!



- Multiple answers like "B or C" constitute own categories
- Opinions are reflected in the most informative way

Main Idea (Couso, Dubois & Sánchez, 2014)

- Extension of the state space $S = \{1, \ldots, c\}$ to $S^* = \mathcal{P}(S) \setminus \{\emptyset\}$
- Interpreting each observation as realisation of a random set (ontic set)
- Basing (precise) analyses on extended state space
- Only adapted state space needs to be considered when interpreting results

Example: Multinomial Logistic Regression (Tutz, 2011)

Classical data situation

Modelling conditional probabilities of response Y for each category $s \in S$, conditional on some covariates $\boldsymbol{x}_i$:

For each category $s \in \{1, \ldots, c-1\}$, $c = |S|$ by

$$P(Y_i = s \,|\boldsymbol{x}_i) \;=\; \frac{\exp(\tilde{\boldsymbol{x}}_i^T \beta_s)}{1 + \sum_{r=1}^{c-1} \exp(\tilde{\boldsymbol{x}}_i^T \beta_r)}$$

and for reference category c by

$$P(Y_i = c \,|\boldsymbol{x}_i) \;=\; \left(1 + \sum_{r=1}^{c-1} \exp(\tilde{\boldsymbol{x}}_i^T \beta_r)\right)^{-1}$$

Example: Multinomial Logistic Regression (Tutz, 2011)

Data under ontic imprecision

Modelling conditional probabilities of response Y^* for each category $s \in S^*$, conditional on some covariates $\boldsymbol{x}_i$:

For each category $s \in \{1, \ldots, m-1\}$, $m = |S^*|$ by

$$P(Y_i^* = s \,|\boldsymbol{x}_i) \;=\; \frac{\exp(\tilde{\boldsymbol{x}}_i^T \beta_{\boldsymbol{s}}^*)}{1 + \sum_{r=1}^{m-1} \exp(\tilde{\boldsymbol{x}}_i^T \beta_{\boldsymbol{r}}^*)}$$

and for reference category m by

$$P(Y_i^* = m \,|\boldsymbol{x}_i) \;=\; \left(1 + \sum_{r=1}^{m-1} \exp(\tilde{\boldsymbol{x}}_i^T \beta_{\boldsymbol{r}}^*)\right)^{-1}$$

Results for German Longitudinal Election Study 2013

Excerpt of regression coefficients regarding First Vote

Coefficient	ontic		classical
	CD	G:S	CD
intercept	0.33	−1.41 **	−0.12
rel.christ	0.37 **	−0.25	0.52 ***
info.tv	−0.02	−0.32	0.25
info.np	−0.12	−1.69 **	0.13

- More regression coefficients
- Even change in sign for some coefficients of *precise* categories
- Some significant *Undecideds'* coefficients

Conclusion & Outlook

- *Just* a change in state space necessary
- Statistical methods do not change, only interpretation
- Presented examples tip of the iceberg

  $\Longrightarrow$ More (complex) models may be applied consequently

Conclusion & Outlook

- *Just* a change in state space necessary
- Statistical methods do not change, only interpretation
- Presented examples tip of the iceberg

  $\Longrightarrow$ More (complex) models may be applied consequently

- Adaptation to ordinal scale
- Ontic imprecision in covariates