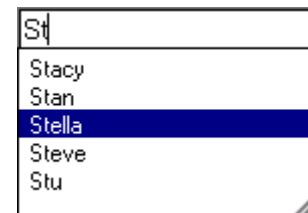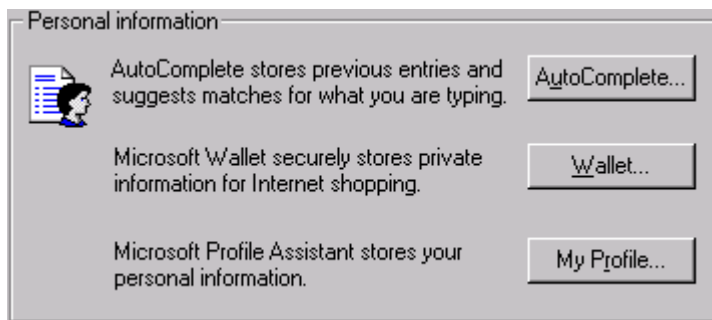# The privacy economics of voluntary over-disclosure in Web forms

Sören Preibusch, **Kat Krol, Alastair R. Beresford**

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

11th WEIS

25th June 2012

Web forms: ubiquitous, versatile, est. 1995

- **Primary mechanism for explicit data collection**

- **Considered a nuisance (time, effort, distraction) Inconclusive user experience advice**

- **Technical easing of form-filling**

Online questionnaires vs. transactional Web forms

Object of assessment and methodology

Users' **behaviour** when providing **optional personal information** on a Web form

- **Field experiment on mTurk: 1500 participants (US)**

- **Web form with 10 optional + 2 check questions**

- **No input checks; participants always paid**

Apparatus, treatments, sample size

About yourself

Please provide some information about yourself. Questions 5 and 6 are mandatory. All other fields are optional. There is no bonus for this HIT.

1. What is your first name?

2. Which city are you in now?

3. What is your favorite color?

4. Do you have any siblings?

5. Which of these questions are mandatory?

6. Do you expect a bonus for this HIT?

7. Is it sunny outside?

8. When did you last spend more than $100?

9. Which browser are you using?

10. Are you in good health?

11. What is your date of birth?

12. Are you a good person?

finish and submit HIT

data req.	compensation	
	$ .25	$ .50
low chk	202	216
high chk, col, sun	209	445
bonus chk, [col, sun]	181	

Revelation ratios by data item and treatment
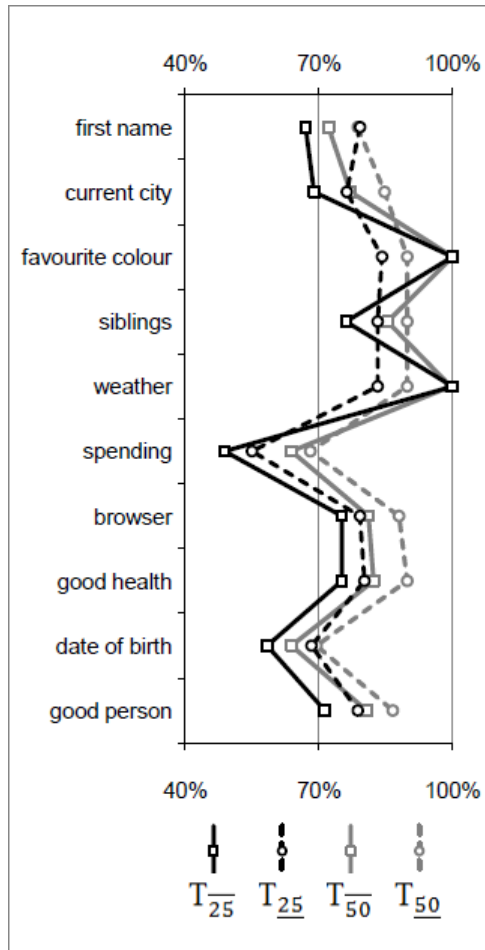
- **Date of birth** disclosed least often: **57%** (partial DOB: **68%**)

- **Weather, favourite colour** disclosed most: **87%**

- Most recent browser used: **66%**

- Browser correctly named: **96%**

Significant prevalence of voluntary over-disclosure

- **Mandatory / optional response drop (p < 0.0001)**

- **Wide-spread voluntary over-disclosure (p < 0.0001)**
  - All vs. none optional fields: 10× as often

- **Significantly over-detailed responses**
  - 6% give weather details
    "No. It's currently cloudy and rainy"

  - 14% give purpose of spending (on top of date)
    "4 days ago getting groceries"

Over-disclosure by accident – but not only!

- **O/D more prevalent when instructions not read**
  - Date of birth: 67% vs. 87%, $p < 0.0001$
  - Good person: 81% vs. 90%, $p = 0.0001$
  - 93% recalled that questions optional

- **Retrospectively personal information**
  - Personal data: 62%
  - Personal and sensitive: 8% (who disclosed DOB, $p < 0.05$)

- **Privacy costs of over-disclosure**

Non-privacy costs of over-disclosure
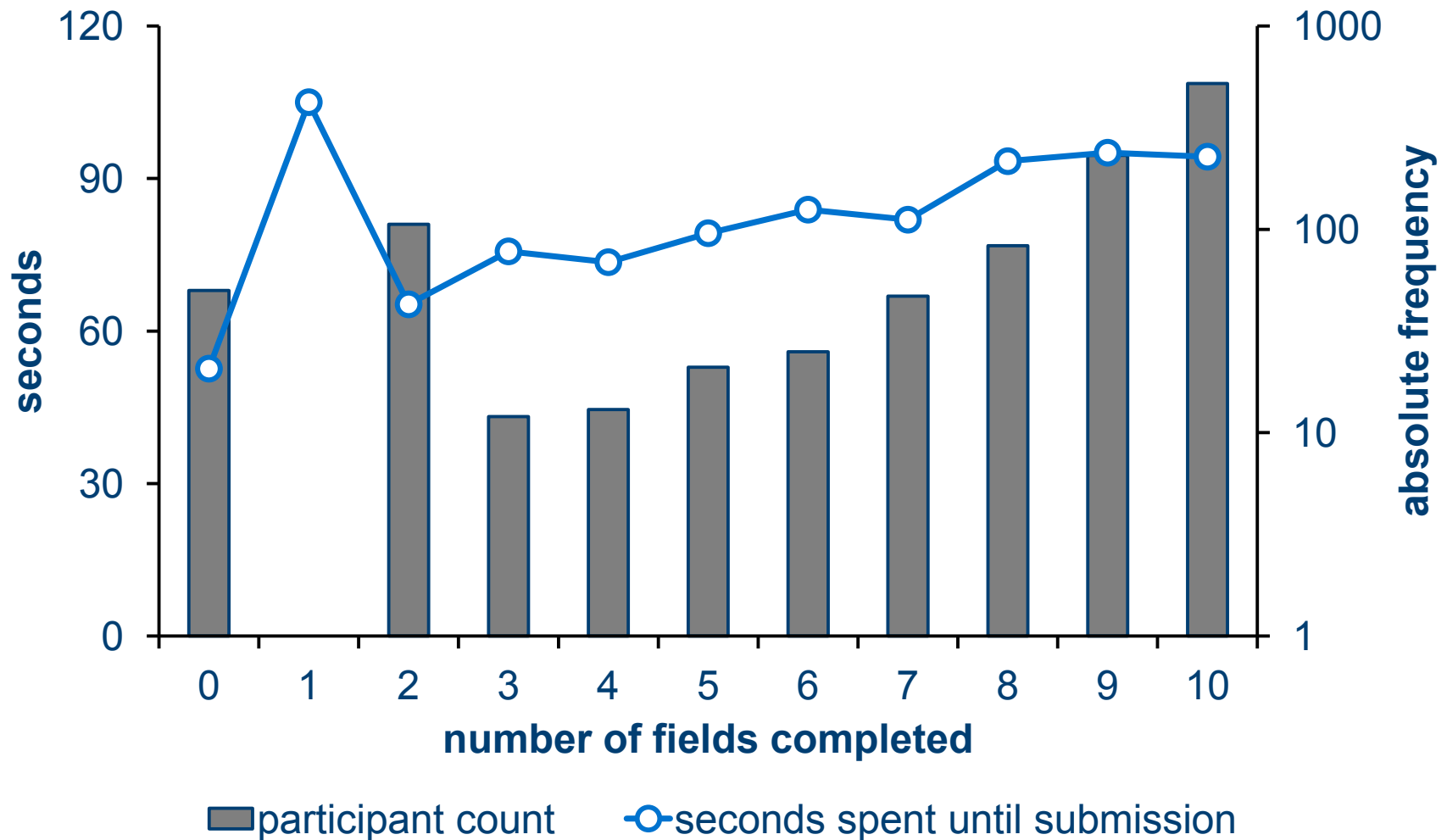
Chilling effects of mandating fields

- **Weather and favourite colour (least sensitive fields) made mandatory: disclosing behaviour reduced**

  - for remaining medium sensitivity
    good person ↘ ($p < 0.04$)

  - for remaining high sensitivity
    date of birth ↘ ($p < 0.02$)

- **Even if mandatory items were provided anyway**

  - average 1.3 fields less provided ($p < 0.0001$)

  - date of birth ↘ ($p < 0.0001$)

Higher base reward may increase disclosure ratio

- **Comparing treatments with high vs. low reward**

- **Effect of higher rewards by data item sensitivity:**
  - Low sensitivity: weather ↗, fav. colour ↗ (p = 0.001)
  - Medium sensitivity: good person ↗ (p = 0.003)
  - High sensitivity: date of birth: no effect (not significant)

- **No moderating effect of reciprocal personality**

Crowding-in of incentivised disclosure

- **Extra $ .25** for disclosing weather and fav. colour

- **Highly effective** for increasing disclosure
  (p < 0.0001; as effective as mandatoriness)

- **Spillover** to non-incentivised data items

  - good person ↗ (p = 0.002)
  - date of birth ↗ (p < 0.001)

Motivation to participate

- **For the money: 54%**

- **Looked easy: 30%**

- **Joy: 15%**

  = "It looked interesting, fun and easy to do"

- **Help research: 8%**

  = "Any help I can be for research, I am glad to do"

- **Interesting: 25%**

- **Articulate opinions: 3%**

  = "my information goes towards creating a change in something"

  = "the opportunity to present an underrepresented demographic (conservatives, mothers) in surveys"

Plausible motives for voluntary over-disclosure

- **By accident** ✓

- **Limit disclosure too costly**

- **Personality**
  - Reciprocity ✗
  - Benevolence
  - Extroversion
  - Completionist ✓

- **Speculating on return**
  - Social capital build-up ✓
  - Monetary bonus ✓
  - Non-monetary returns (e.g., personalisation)
  - Infrastructure improvements ✓
  - Opinion shaping ✓

- Web users incur costs from highly prevalent, voluntary over-disclosure: time, effort, privacy loss.

- Incentives create positive spillover towards higher disclosure.

- Mandating some fields reduces voluntary disclosure for the remaining.