The Prototype of Privacy: Analyzing Privacy Discourse through its Features

Mina Vasalou (University of Bath)
Alastair Gill (University of Surrey)
Fadhila Mazanderani, Chrysanthi Papoutsi,
Anne-Marie Oostveen (University of Oxford)
Sacha Brostoff (University College London)

Theoretical challenges in privacy

- There is a lack of a clear definition or consensus on privacy.
- Theoretical disputes on what ‘criteria’ belong to the concept of privacy
- Should it be studied as a human right or as an interest motivated by an individual’s goals?

Methodological challenges in privacy

- Asking about privacy → primes privacy concerns
- We need unobtrusive methods for measuring privacy
- For qualitative research, linguistic analysis is a possible solution
- Using a theory driven approach we develop a privacy dictionary that can show linguistic differences in people's privacy discourse

Prototype Theory

- Evolved from Wittgenstein's family resemblance
- Prototype theory proposes that the boundaries of concepts are fuzzy and there are better and worse exemplars
- More prototypical members have more features in common (e.g. is an apple a good member of fruit? Is an apple recognized faster than an olive as a member of fruit?)
- Methodologies have been established in previous prototype studies on social concepts such as forgiveness, love, gratitude

Methodology

- Obtained a list of 82 features from 146 participants
- 118 Participants rated how central or important the features were to privacy (central, intermediate, peripheral)
- 62 participants rated the extend to which created vignettes (with central, intermediate, and peripheral features) were representative examples of privacy

Privacy is organized around a prototype

- 82 features
- Reliably rated on centrality
- Vignettes with central features were judged as better exemplars of privacy than vignettes with peripheral features
- In a memory task, the prototype was first activated. Participants falsely recognized central non-presented items.

Dictionary was built on the features

1. 82 features were broken down into seed words e.g. ‘securing my personal information’
2. Synonyms and antonyms of seed words were identified (total of 604 words)
3. Collected a dataset of interview transcripts from 7 diverse offline and online privacy-sensitive contexts.

Analysis of interviews

- Raters marked up areas of text talking about privacy
- The privacy texts (50.000 words) were compared to the non-privacy text (200.000 words).
- We omitted 373 words from the 604:
  - …that were not mentioned in the privacy texts
  - …that were not used as anticipated e.g. company

Still to do…

- Use existing content and linguistic analysis tools to examine linguistic differences between privacy and non-privacy text (at word, content, syntax and semantic levels)
- Build a privacy dictionary tool which will be tested on new data. The results will be compared to human raters' judgement.
- The automated privacy dictionary will assist researchers in detecting privacy relevant discourse and will maximize the richness of qualitative analysis.