“I need a better description”: An Investigation Into User Expectations For Differential Privacy

Gabriel Kaptchuk (Boston University)
joint work with
Rachel Cummings (Columbia University)
Elissa M Redmiles (Max Plank Institute for Software Systems)
Differential privacy [DMNS ‘06]
Differential privacy [DMNS ‘06]

• Differential privacy is deployed in practice by major tech companies and government organizations:

  ![Logos of Google, Apple, Uber, and Facebook]

• How should these organizations explain differential privacy to their users?
I want to, but I’m concerned about privacy!

You should sign up for my product. It does cool things!

We use differential privacy!

What’s that??

Good Company is the logo of a Ski company – No endorsement implied
What happens on your iPhone, stays on your iPhone.
apple.com/privacy
Research questions

Does DP “work” for users?

How do users understand DP when they encounter it “in-the-wild?”
Research questions

(RQ1) Do users care about the type of protections provided by DP?
(RQ2) Are users more willing to share their data with reduce information disclosure risks?
(RQ3) How does the way DP is described impact users’ expectations of protection against information disclosures?
(RQ4) How does the way DP is described impact users’ willingness to share their data?
Research plan / Outline

• Vignette-based surveys to elicit preferences/perceptions (n=2,424)

• To address RQ1 and RQ2,
  • Present information-sharing scenario, and query privacy concerns
  • Set privacy expectations for those concerns, and query willingness to share data
  • Measures how users’ privacy concerns align with the protections provided by DP

• To address RQ3 and RQ4,
  • Collect descriptions of differential privacy
  • Present information-sharing scenario and a DP description
  • Query privacy expectations and willingness to share data
  • Measures how accurately and effectively DP descriptions set user expectations
For information of survey methods, please consult the paper.
## Information disclosure possibilities

<table>
<thead>
<tr>
<th>Name</th>
<th>Expectation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hack</td>
<td>A criminal or foreign government that hacks the organization could learn my data</td>
</tr>
<tr>
<td>Law Enforcement</td>
<td>Law enforcement could access my data with a court order</td>
</tr>
<tr>
<td>Organization</td>
<td>My friend will learn my data/ Organization will store data</td>
</tr>
<tr>
<td>Data Analyst</td>
<td>A data analyst working for the organization could learn my data</td>
</tr>
<tr>
<td>Graphs</td>
<td>Graphs or informational charts created using the collected information could reveal my data</td>
</tr>
<tr>
<td>Sharing</td>
<td>Data that the organization shares with other organizations could reveal my data</td>
</tr>
</tbody>
</table>
RQ1: What information disclosures concern users?

- Share: 60.3%
- Organization: 55.3%
- Hack: 52.1%
- Data Analyst: 43.5%
- Graphs: 40.8%
- Law Enforcement: 39.6%

Proportion of Respondents (n=1216)
RQ2: How does the probability of information disclosures affect data sharing?

- Logistic regression model for respondents who cared about disclosure
- Odds Ratio measures change in odds of sharing data, compared against High Risk condition (OR>1 $\implies$ increased sharing), CI: 95%
- Read as: “decreased risk of <disclosure> increased chances of sharing by X”

<table>
<thead>
<tr>
<th>Variable</th>
<th>Hack</th>
<th>Law Enforcement</th>
<th>Organization</th>
<th>Data Analyst</th>
<th>Graphs</th>
<th>Share</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR/CI</td>
<td>p-value</td>
<td>OR/CI</td>
<td>p-value</td>
<td>OR/CI</td>
<td>p-value</td>
</tr>
<tr>
<td>Low Risk</td>
<td>1.91</td>
<td>&lt; 0.01**</td>
<td>0.87</td>
<td>0.55</td>
<td>1.63</td>
<td>0.01*</td>
</tr>
<tr>
<td></td>
<td>[1.27, 2.88]</td>
<td></td>
<td>[0.54, 1.39]</td>
<td></td>
<td>[1.12, 2.38]</td>
<td></td>
</tr>
<tr>
<td>No Risk</td>
<td>2.97</td>
<td>&lt; 0.01***</td>
<td>2.07</td>
<td>&lt; 0.01**</td>
<td>1.61</td>
<td>0.01*</td>
</tr>
<tr>
<td></td>
<td>[1.98, 4.49]</td>
<td></td>
<td>[1.32, 3.27]</td>
<td></td>
<td>[1.1, 2.35]</td>
<td></td>
</tr>
</tbody>
</table>

- OR > 1 $\implies$ increased sharing, CI: 95%
- Read as: “decreased risk of <disclosure> increased chances of sharing by X”
Survey 2

Descriptions → Expectations → Willingness to Share
We use differential privacy!

“differential privacy,” the new **gold standard in data privacy** protection.

Differential privacy works by algorithmically **scrambling individual user data** so that it cannot be traced back.

When a differential privacy algorithm is applied to a data set, those **links get blurred**, and bits of data can no longer be traced to their source.

“In ideal implementations, this risk remains close to zero, guaranteeing... virtually no adverse effect on them from an informational standpoint.”

“differential privacy,” which alters the numbers but **does not change core findings** to protect the identities of individual respondents.

In short, differential privacy **allows general statistical analysis** without revealing information about a particular individual in the data.
Differential Privacy Descriptions Gathered

<table>
<thead>
<tr>
<th>Industry</th>
<th>Press</th>
<th>Academic</th>
</tr>
</thead>
<tbody>
<tr>
<td>37</td>
<td>30</td>
<td>10</td>
</tr>
</tbody>
</table>
Resulting Descriptions:

1. Unsubstantial
2. DP Techniques
3. DP enables analysis
4. DP is widely trusted
5. DP reduces user risk
6. Technical description of DP
Differential privacy...

is the gold standard in data privacy and protection and is widely recognized as the strongest guarantee of privacy available.

4. DP is widely trusted 5. DP reduces user risk 6. Technical description of DP
Differential privacy...

injects statistical noise into collected data in a way that protects privacy without significantly changing conclusions.
Differential privacy...

allows analysts to learn useful information from large amounts of data without compromising an individual's privacy.
Differential privacy...

is a novel mathematical technique to preserve privacy which is used by companies like Apple and Uber.
Differential privacy...

protects a user’s identity and the specifics of their data, meaning individuals incur almost no risk by joining the dataset.
Differential privacy...

ensures that the removal or addition of a single database item does not (substantially) affect the outcome of any analysis. It follows that no risk is incurred by joining the database, providing a mathematically rigorous means of coping with the fact that distributional information may be disclosive. [Dwork08]
**RQ3: How do differential privacy descriptions affect privacy expectations?**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Hack</th>
<th>OR/CI</th>
<th>p-value</th>
<th>Law Enforcement</th>
<th>OR/CI</th>
<th>p-value</th>
<th>Organization</th>
<th>OR/CI</th>
<th>p-value</th>
<th>Data Analyst</th>
<th>OR/CI</th>
<th>p-value</th>
<th>Graphs</th>
<th>OR/CI</th>
<th>p-value</th>
<th>Share</th>
<th>OR/CI</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description:</td>
<td>1.94</td>
<td>0.01*</td>
<td></td>
<td>1.10</td>
<td>0.72</td>
<td></td>
<td>1.13</td>
<td>0.59</td>
<td></td>
<td>1.71</td>
<td>0.1</td>
<td></td>
<td>1.64</td>
<td>0.05*</td>
<td></td>
<td>1.68</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>Unsubstantial</td>
<td>[1.16, 3.29]</td>
<td>[0.65, 1.86]</td>
<td></td>
<td>[0.73, 1.75]</td>
<td>[0.92, 3.27]</td>
<td></td>
<td>[1.01, 2.67]</td>
<td>[1.01, 2.67]</td>
<td></td>
<td>[1.09, 2.87]</td>
<td>[0.99, 2.88]</td>
<td></td>
<td>[1.33, 3.77]</td>
<td>[1.33, 3.77]</td>
<td></td>
<td>[1.33, 3.77]</td>
<td>[1.33, 3.77]</td>
<td></td>
</tr>
<tr>
<td>Description:</td>
<td>1.96</td>
<td>0.01*</td>
<td></td>
<td>1.21</td>
<td>0.47</td>
<td></td>
<td>1.43</td>
<td>0.1</td>
<td></td>
<td>2.40 &lt; 0.01**</td>
<td>2.15 &lt; 0.01**</td>
<td></td>
<td>2.22 &lt; 0.01**</td>
<td>[1.34, 3.5]</td>
<td></td>
<td>[1.69]</td>
<td>[1.29]</td>
<td></td>
</tr>
<tr>
<td>Techniques</td>
<td>[1.17, 3.33]</td>
<td>[0.72, 2.03]</td>
<td></td>
<td>[0.93, 2.22]</td>
<td>[1.33, 4.5]</td>
<td></td>
<td>[1.34, 4.5]</td>
<td>[1.34, 4.5]</td>
<td></td>
<td>[1.09, 2.87]</td>
<td>[1.34, 4.5]</td>
<td></td>
<td>[1.69]</td>
<td>[1.29]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Description:</td>
<td>1.60</td>
<td>0.08</td>
<td></td>
<td>1.05</td>
<td>0.84</td>
<td></td>
<td>1.40</td>
<td>0.13</td>
<td></td>
<td>2.06 &lt; 0.01**</td>
<td>1.76 &lt; 0.01**</td>
<td></td>
<td>2.22 &lt; 0.01**</td>
<td>[1.34, 4.5]</td>
<td></td>
<td>[1.69]</td>
<td>[1.34, 4.5]</td>
<td></td>
</tr>
<tr>
<td>Enables</td>
<td>[0.95, 2.73]</td>
<td>[0.63, 1.77]</td>
<td></td>
<td>[0.91, 2.16]</td>
<td>[1.13, 3.88]</td>
<td></td>
<td>[1.34, 4.5]</td>
<td>[1.34, 4.5]</td>
<td></td>
<td>[1.09, 2.87]</td>
<td>[1.34, 4.5]</td>
<td></td>
<td>[1.69]</td>
<td>[1.34, 4.5]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Description:</td>
<td>1.86</td>
<td>0.02*</td>
<td></td>
<td>1.04</td>
<td>0.89</td>
<td></td>
<td>1.43</td>
<td>0.11</td>
<td></td>
<td>1.99 &lt; 0.01**</td>
<td>1.38 &lt; 0.01**</td>
<td></td>
<td>2.27 &lt; 0.01**</td>
<td>[1.34, 4.5]</td>
<td></td>
<td>[1.69]</td>
<td>[1.34, 4.5]</td>
<td></td>
</tr>
<tr>
<td>Trust</td>
<td>[1.11, 3.17]</td>
<td>[0.61, 1.76]</td>
<td></td>
<td>[0.92, 2.22]</td>
<td>[1.08, 3.78]</td>
<td></td>
<td>[1.08, 3.78]</td>
<td>[1.08, 3.78]</td>
<td></td>
<td>[0.84, 2.28]</td>
<td>[0.84, 2.28]</td>
<td></td>
<td>[1.19]</td>
<td>[1.34, 4.5]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Description:</td>
<td>2.58</td>
<td>&lt; 0.01***</td>
<td></td>
<td>1.86</td>
<td>&lt; 0.01***</td>
<td></td>
<td>1.43 &lt; 0.01**</td>
<td>0.1</td>
<td></td>
<td>2.46 &lt; 0.01**</td>
<td>2.40 &lt; 0.01***</td>
<td></td>
<td>2.27 &lt; 0.01***</td>
<td>[1.34, 4.5]</td>
<td></td>
<td>[1.69]</td>
<td>[1.34, 4.5]</td>
<td></td>
</tr>
<tr>
<td>Risk</td>
<td>[1.57, 4.33]</td>
<td>[1.15, 3.05]</td>
<td></td>
<td>[0.93, 2.2]</td>
<td>[1.37, 4.59]</td>
<td></td>
<td>[1.37, 4.59]</td>
<td>[1.37, 4.59]</td>
<td></td>
<td>[1.5, 3.88]</td>
<td>[1.37, 4.59]</td>
<td></td>
<td>[1.69]</td>
<td>[1.34, 4.5]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Description:</td>
<td>1.56</td>
<td>0.1</td>
<td></td>
<td>1.02</td>
<td>0.95</td>
<td></td>
<td>1.38</td>
<td>0.15</td>
<td></td>
<td>2.30 &lt; 0.01**</td>
<td>1.70 &lt; 0.01**</td>
<td></td>
<td>1.90 &lt; 0.01**</td>
<td>[1.04, 2.79]</td>
<td></td>
<td>[1.69]</td>
<td>[1.34, 4.5]</td>
<td></td>
</tr>
<tr>
<td>Technical</td>
<td>[0.92, 2.69]</td>
<td>[0.6, 1.73]</td>
<td></td>
<td>[0.89, 2.14]</td>
<td>[1.26, 4.33]</td>
<td></td>
<td>[1.26, 4.33]</td>
<td>[1.26, 4.33]</td>
<td></td>
<td>[1.04, 2.79]</td>
<td>[1.26, 4.33]</td>
<td></td>
<td>[1.69]</td>
<td>[1.34, 4.5]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

• Read as: “Describing DP with <description> increases privacy expectations for <disclosure> by multiplicative factor of X”
Central Model

Central Aggregator

Private Output

Local Model

Central Aggregator

Private Output
Ground truth* vulnerabilities – Local vs central DP

<table>
<thead>
<tr>
<th>Disclosure</th>
<th>Local</th>
<th>Central</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hack</td>
<td>False</td>
<td>True</td>
</tr>
<tr>
<td>Law Enforcement</td>
<td>False</td>
<td>True</td>
</tr>
<tr>
<td>Organization</td>
<td>False</td>
<td>True</td>
</tr>
<tr>
<td>Data Analyst</td>
<td>False</td>
<td>True</td>
</tr>
<tr>
<td>Graphs</td>
<td>False</td>
<td>False</td>
</tr>
<tr>
<td>Sharing</td>
<td>False</td>
<td>True</td>
</tr>
</tbody>
</table>

*For “typical” implementation. Actual ground truth will depend on implementation details, including privacy parameters.
Correctness of expectations – Local vs central DP
RQ4: How do descriptions of DP affect sharing?

<table>
<thead>
<tr>
<th>Variable</th>
<th>Odds Ratio</th>
<th>CI</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description: Unsubstantial</td>
<td>1.22</td>
<td>[0.79, 1.88]</td>
<td>0.37</td>
</tr>
<tr>
<td>Description: Techniques</td>
<td>0.96</td>
<td>[0.62, 1.47]</td>
<td>0.83</td>
</tr>
<tr>
<td>Description: Enables</td>
<td>1.48</td>
<td>[0.96, 2.29]</td>
<td>0.08</td>
</tr>
<tr>
<td>Description: Trust</td>
<td>1.08</td>
<td>[0.7, 1.67]</td>
<td>0.72</td>
</tr>
<tr>
<td>Description: Risk</td>
<td>1.37</td>
<td>[0.89, 2.12]</td>
<td>0.15</td>
</tr>
<tr>
<td>Description: Technical</td>
<td>0.94</td>
<td>[0.61, 1.45]</td>
<td>0.77</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Odds Ratio</th>
<th>CI</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salary Scenario</td>
<td>1.67</td>
<td>[1.32, 2.1]</td>
<td>&lt;0.01***</td>
</tr>
<tr>
<td>Internet Score</td>
<td>1.09</td>
<td>[0.95, 1.25]</td>
<td>0.2</td>
</tr>
</tbody>
</table>

- DP descriptions have no significant effect
- Information sharing scenario does effect sharing intention
(RQ1) Users care about DP protections
(RQ2) Users respond to better protections
(RQ3) DP descriptions (haphazardly) raise privacy expectations

(RQ4) Descriptions of DP do not increase willingness to share
“Differential Privacy allows analysts to learn useful information from large amounts of data without compromising an individual's privacy.”

I’m more likely to share my information if it’s protected from hackers and law enforcement.

Raising some expectations (eg. Graphs and data analyst)

No increased sharing!
RQ1: What information disclosures concern users?

- **Share**: 60.3%
- **Organization**: 55.3%
- **Hack**: 52.1%
- **Data Analyst**: 43.5%
- **Graphs**: 40.8%
- **Law Enforcement**: 39.6%

Proportion of Respondents (n=1216)
Misalignment of User Concern and Description Effects

Legend:
- Hack
- Law Enforcement
- Organization
- Data Analyst
- Graphs
- Share

No Increased Sharing
- No Overlap
  - User Concerns: 
  - Description Effects: 

Some Increased Sharing
- Partial Overlap
  - User Concerns: 
  - Description Effects: 

Significantly Increased Sharing
- Full Overlap
  - User Concerns: 
  - Description Effects: 

Description: Enables
Description: Risk
Description: Techniques
(RQ1) Users care about DP protections
(RQ2) Users respond to better protections
(RQ3) DP descriptions (haphazardly) raise privacy expectations
(RQ4) Descriptions of DP do not (directly) increase willingness to share
Alignment between descriptions and preferences seems crucial
Thanks!

Gabriel Kaptchuk, Boston University

Big thanks to my collaborators Rachel Cummings and Elissa Redmiles