Lessons Learned From Evaluating Eight Password Nudges in the Wild

Karen Renaud, Verena Zimmermann, Joseph Maguire & Steve Draper

Supported by:

Abertay University

University of Glasgow

Technische Universität Darmstadt

Bundesministerium für Bildung und Forschung

Hessisches Ministerium für Wissenschaft und Kunst
Research Question

Users often create weak passwords and hackers are easily able to compromise accounts. [1,2]

Can we find a way to “nudge” people towards better passwords?
The Concept of Nudging

A nudge attempts to influence people towards a wiser option by manipulating the choice architecture surrounding the behavior to encourage wiser choices.
Nudging has Worked

Examples from the Behavioural Sciences:

• Improving tax repayment percentages [5]
• Reducing speeding [6]
• Opt-Out vs Opt-In for organ donations [7]
Nudges and IT Security?

• People can be successfully nudged towards a secure WiFi [8]
• Nudging has helped to steer people away from apps that request too many permissions [9]
• Password Strength Meters? → Inconclusive results
  • Ur et al. (2012) found a positive impact
Method: Apparatus and Procedure

- Two sequential studies running for one academic year each
- Use of a university web application (grades, feedback, coursework deadlines etc.) → Important and frequently-used password
- Display of visual nudges on registration page of web application
- Random assignment of students to either control or one of five nudge conditions
- Informed consent and possibility to opt out

- 497 participants in study 1, 779 participants in study 2
- Mainly Computer Science students
**Method: Study Design**

<table>
<thead>
<tr>
<th>Year 1</th>
<th>Year 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>IV: Nudge Condition</td>
<td>IV: Nudge Condition</td>
</tr>
<tr>
<td>N0 Control</td>
<td>N0 Control</td>
</tr>
<tr>
<td>N1 Framing Effect</td>
<td>N2 Expectation Effect</td>
</tr>
<tr>
<td>N2 Expectation Effect</td>
<td>N3 In-Group Effect</td>
</tr>
<tr>
<td>N3 In-Group Effect</td>
<td>N6 Social Norms</td>
</tr>
<tr>
<td>N4 Expectation Effect + Strength</td>
<td>N7 Expectation Effect + Reflection</td>
</tr>
<tr>
<td>N5 In-Group Effect + Strength</td>
<td>N8 In-Group Effect + Effect</td>
</tr>
<tr>
<td>DV: Password length</td>
<td>DV: Password length</td>
</tr>
<tr>
<td>Password strength</td>
<td>Password strength</td>
</tr>
</tbody>
</table>
Method: DV

- Password length: measured in number of characters
- Password strength: measured by score metric provided by strength estimator zxcvbn [10]
  - 0 - number of guesses $< 10^2$
  - 1 - number of guesses $< 10^4$
  - 2 - number of guesses $< 10^6$
  - 3 - number of guesses $< 10^8$
  - 4 - number of guesses above
Year 1
Most systems prompt for a password with the word “Password”

The second most common password is also “password”

What would happen if we changed the word to “secret”?
Year 1: N2/N4 - Expectation Effect (Strength)
Year 1: N3/N5 – In-Group Effect (Strength)
Study 1: Results

Analysis:

- Password strength – ordinal scale, Password length – not normally distributed
- Use of non-parametric Mann-Whitney-U tests and Benjamini-Hochberg correction
- Pairwise comparisons: Control against Nudge Conditions

Results:

- No indication for the framing effect (“Secret”)
- No significant differences between control and nudge conditions found
- Small effect sizes Cliff’s Delta between 0.02 and 0.17
What next?

Sunstein (2017) suggests:

1. Give up, if you have reason to believe that the user knows best
2. Try different nudges
3. Offer an Economic Incentive [12]
Year 2
As a student, how strong do you think this password is?

- Very Weak
- Weak
- OK
- Strong
- Very Strong
- Unsure
As a SOCS student, how strong do you think this password is?

- Very Weak
- Weak
- OK
- Strong
- Very Strong
- Unsure
Year 2: N6 – Social Norms
Year 2: Results

• Similar procedure as in study 1
• No significant differences between control group and Nudge Conditions found

Questions:
• Why is this?
• What do the results mean?
• What shall we do next?
Discussion and Reflection
Methodological Issues

The strength metric

- Ordinal scale required use of nonparametric tests
  - Test power of parametric tests is slightly higher (up to 2%)
- Artificial categorization led to loss of variance and information
  - E.g. passwords that required 1100 vs. 9900 guesses to be broken would both be assigned score 2 (number of guesses between $10^4$ and $10^6$)

→ Differences existent, but not detectable due to choice of DV and analysis?
User Issues

Authentication is complex
• Authentication is a secondary task
• User has primary task, goals and needs → not fully considered?

Password strength perceptions
• Password strength perceptions can differ from actual password strength [13]
• Nudges only indicated that passwords should be stronger → but how?

Password Reuse
• Passwords might have been reused instead of created [14] → creation process not influenced by nudge?
Lessons learned

The strength metric

• Different choice of dependent variable, search for alternative metrics

The User

• User context should be fully considered
• Can password reuse be prevented?
• Beyond pure nudging:
  • Offer a benefit for stronger passwords?
  • Provide feedback/instruction on how to increase password strength?
Are there any Questions?

Contact
Verena Zimmermann
Mail: zimmermann@psychologie.tu-darmstadt.de
LITERATURE
Literature


Literature


